PROCEEDINGS

OF THE

American Philosophical Society

Held at Philadelphia for Promoting Useful Knowledge

VOLUME 103, NUMBER 3 JUNE 15, 1959

CONTENTS

The First World War, 1914–1918. BERNADOTTE E. SCHMITT 321

A Stratigraphic Excavation at El Risco, Mexico. WILLIAM J. MAYER-OAKES 332

Roman Men in Greek Himation (Roman Palliati). A Contribution to the History of Copying. MARGARETE BIEBER 374

Improvisation in the Middle English Romance. ALBERT C. BAUGH 418

The Challenge of the Psychopharmaceuticals. NATHAN S. KLINE 455

The Common Law and Codification in Jacksonian America. PERRY MILLER 463

Bernard Shaw, Sidney Webb, and the Theory of Fabian Socialism. GEORGE J. STIGLER 469

Dr. John Evans, U. S. Geologist to the Oregon and Washington Territories. ERWIN F. LANGE 476

An Axiomatization of General Relativity. RICHARD A. MOULD 485

THE AMERICAN PHILOSOPHICAL SOCIETY

INDEPENDENCE SQUARE

PHILADELPHIA 6

1959

PUBLICATIONS

OF

The American Philosophical Society

The publications of the American Philosophical Society consist of PROCEEDINGS, TRANSACTIONS, MEMOIRS, and YEAR BOOK.

THE PROCEEDINGS contains papers which have been read before the Society in addition to other papers which have been accepted for publication by the Committee on Publications. In accordance with the present policy one volume is issued each year, consisting of six bimonthly numbers, and the price is $5.00 net per volume. Current price for individual numbers is $1.00.

THE TRANSACTIONS, the oldest scholarly journal in America, was started in 1769 and is quarto size. In accordance with the present policy each annual volume is a collection of monographs each issued as a part. The annual subscription price is $6.00 net per volume; for both the PROCEEDINGS and the TRANSACTIONS $10.00.

Each volume of the MEMOIRS is published as a book. The titles cover the various fields of learning; most of the recent volumes have been historical. The price of each volume is determined by its size and character.

The YEAR BOOK is of considerable interest to scholars because of the reports on grants for research and to libraries for this reason and because of the section dealing with the acquisitions of the Library. In addition it contains the Charter and Laws, and lists of present and former members, and reports of committees and meetings. The YEAR BOOK is published about April 1 for the preceding calendar year. The price is $1.50.

An author desiring to submit a manuscript for publication should send it to the Editor, William J. Robbins, American Philosophical Society, 104 South Fifth Street, Philadelphia 6, Pa.

THE FIRST WORLD WAR, 1914–1918

BERNADOTTE E. SCHMITT

Andrew MacLeish Distinguished Service Professor Emeritus of Modern History, University of Chicago

(Read November 14, 1958)

FORTY years ago last Tuesday, on November 11, 1918, the First World War was brought to an end by the armistice signed with Germany. The First World War is now almost forgotten. Yet the terrible age in which we live derives in clear succession from the struggle of 1914–1918. If the present "cold war" is the aftermath of the Second World War, that conflict arose from the refusal of Germany to accept the defeat of 1918 as conclusive, while, furthermore, the Communist menace which overshadows our lives dates from 1917. It is therefore appropriate to take a look back to the years 1914–1918, which were perhaps more tortured to those who lived through them than were the even more rugged years of 1939–1945.

The Second World War had long been expected and men's minds were hardened to it in advance, whereas the catastrophe of 1914 came as an unbelievable surprise to a generation which had not experienced a general European war for a hundred years.

In retrospect, how wonderful the years before 1914 appear! Europe seemed politically stable, for in spite of periodic crises during the previous decade, the will to peace had triumphed. In western Europe, governments were fairly democratic, being responsible to parliaments elected by manhood suffrage. In central Europe, constitutional government prevailed, and even in Russia autocracy was tempered by the Duma. The restless Balkans, and Turkey as well, had their parliaments. Kings were the rule—and generally popular; republics were the exception. By and large the rule of law was recognized; arbitrary tyranny was largely a thing of the past.

Economically, Europe was generally prosperous, thanks to the universal adoption of modern technology and the widespread importation of food from the New World. The gold standard of money prevailed universally and currencies were interchangeable, without restriction of amount. A pound sterling was worth $4.86, a French or Belgian or Swiss franc or an Italian lira was worth 19.5 cents, a German mark was worth 23.8 cents—and the value was constant. Raw materials were brought freely from the ends of the earth and made up into goods which were sold all over the world. A German could buy or sell as freely in the colonies of the British Empire as an Englishman. European control of Africa and Asia, which was to no small degree the secret of European well-being, had not yet been challenged seriously. Perhaps the best index of a peaceful Europe and an orderly world was that travel was almost completely free, passports being necessary only for Russia and Turkey. Free also was the press. Even in Germany and Russia, the least liberal countries in Europe, censorship was intermittent and erratic, and newspapers frequently indulged in savage criticism of the government. Even the All-Highest, the Kaiser himself, was not spared.

Actually, this picture of a golden age, of an ordered, stable and pacific Europe, was deceptive. In greater or less degree, social unrest was widespread and growing more and more active. The working classes were persuaded that they were not receiving a fair share of the ever-increasing wealth produced by modern industry and in every country had formed parties which advocated Socialism as the cure for all the ills of society. Although the Socialist parties proposed to carry out their programs by constitutional methods, they inspired much fear among the governing and middle classes, not only because their economic and social demands were unwelcome, but also because their loyalty in time of war was questioned.

A second circumstance boding ill for the future was the situation in the Balkan Peninsula, where war and revolution had been endemic for a century. To be sure, the great upheaval of 1912–1913, in which the small Balkan states had driven the Turks practically out of Europe, had been accomplished, contrary to expectations, without a general European war, and the restraint displayed by the Great Powers encouraged the hope that in the final analysis the Powers would

compromise rather than fight. But the Balkan peace settlement of 1913 had left many resentments, and the possibility of an explosion could not be ruled out.

Next, the peace of Europe was threatened by the restlessness of Germany. Germany was unquestionably the most advanced nation in Europe technologically, and it was expanding its commerce and increasing its population at a rate which alarmed its neighbors. Millions of Germans were obsessed by the idea that Germany did not enjoy its proper "place in the sun," as the phrase went, for the German colonies were inconsequential when compared with the vast empires of Britain and France, or even those of Belgium and Portugal, and they were not bashful about proclaiming loudly and often that a new deal was imperative. The German army was the strongest in the world; the German navy was being rapidly expanded to challenge the British fleet. Individual Germans talked much about the necessity or grandeur of war, while the official policy of the German government too often took the form of threat and sabre-rattling, as in 1905 and again in 1911; while the Kaiser contributed not a little to uneasiness by frequent and flamboyant speeches. It is perhaps not easy for people in 1958 to remember that fifty years ago Germany seemed to offer the same threat to peace that the Soviet Union does today to our security. William II was as incalculable as Nikita Khruschchev.

Finally, Europe was criss-crossed with alliances. Germany, Austria-Hungary (how many people today understand what Austria-Hungary was?), and Italy formed a Triple Alliance. This was opposed by a Dual Alliance of France and Russia, which became a Triple Entente when Great Britain patched up long-standing quarrels with those powers and joined with them diplomatically. Originally both alliances had been defensive, becoming operative only in the event of attack by other powers, but time and circumstance had modified them. Germany had promised to support Austria if that power attacked Serbia, and later editions of the Triple Alliance recognized Italian aspirations in Africa and even in Europe. France and Russia came to agree that the balance of power in Europe must be maintained, which might involve territorial changes that might or might not be effected peacefully. Although Great Britain was not formally committed to France, nevertheless its interests seemed to require that it should not let France be overpowered by Germany.

In 1914 the two groups stood face to face, and the relations between the members of each group were more intimate than they had ever been. The military and naval power of the two groups was so evenly balanced that during the prolonged crisis of the Balkan wars, from October 1912 to August 1913 neither side dared challenge the other to a showdown. But, if, for any reason, one Power did move, that is, take military action, then—such were the interlocking provisions of the alliances—all were bound to follow. For a generation the system of alliances had kept the peace, but in 1914 its workings produced a general war out of a local dispute between Austria-Hungary and Serbia.

The Austro-Serbian dispute was not beyond solution. At any rate there were men in both Austria-Hungary and Serbia who thought a peaceful solution possible. But the ruling group in Austria did not desire such a solution. Rather it had for years sought an excuse for war, because these men believed that only by war could the anomalous state Austria-Hungary be preserved and their own privileged position safeguarded. Fate provided this excuse when the Archduke Francis Ferdinand, the heir to the Habsburg throne, was murdered at Sarajevo on June 28, 1914. There is still much mystery about that crime. It is not clear how much the Serbian government knew of the plot in advance or what it did or did not do to stop it. On the fatal day, when a first attempt to kill the Archduke had failed, the plans were changed, but the chauffeur of the royal car was not informed of the change—why, is not known. Had he been properly instructed, the assassins would not have had a second chance to shoot. Thus, chance was responsible for providing the long-desired excuse. Had there been no murder at Sarajevo, there would probably not have been a European War in 1914.

Hardly had the war begun when the belligerent governments issued collections of diplomatic papers—"books" they were called, of various hues of color—designed to prove that the other side was responsible for the war. These "books" excited a tremendous controversy among scholars, publicists, and politicians, but the more they debated, the more strongly did each side become convinced of its own innocence and the guilt of the enemy. At the end of the war, revolutions in

Russia, Austria, and Germany led to the opening of archives of the fallen regimes, and presently it became evident that the documents issued in 1914 had been "edited," that is, inconvenient passages had been omitted or changed, and sometimes new documents had been invented. Ultimately, the authentic texts of the 1914 papers were released and thousands of them are now available. Although some points still remain obscure, the main lines are clear enough.

(1) When Austria-Hungary decided to take action against Serbia for the crime at Sarajevo, it asked Germany for support, for it contemplated military action and the partition of Serbia—steps which might well bring Russia on the scene. The German government approved the Austrian program and urged its immediate execution. The Germans calculated that since a royal personage had been murdered, the Tsar of Russia would probably not go to the help of Serbia; but if he did and war resulted, Germany and Austria would be able to defeat Russia and France. The German Emperor and his chancellor assumed that Britain would not take part in such a war. Why they thought this remains a mystery. For many years the German ambassadors in London had reported that if Germany attacked France, Britain would come into the war. Furthermore, the German plans involved violation of the neutrality of Belgium, and Britain was known to be peculiarly sensitive about Belgium. Probably the explanation of this German miscalculation is that the German general staff did not worry about British intervention, for it believed that intervention would come too late to be effective; or, if the British army did get to Belgium, it would be easily annihilated. It is also established that when the German chancellor, Bethmann Hollweg, accepted the Austrian plan, he expressed the opinion that it would be better to have war then rather than at a later date, when the Entente powers would be stronger. Therefore, it can be said that, while there is no evidence that the German government deliberately plotted war for the summer of 1914, when the possibility of war was opened up by the Austrian program, this prospect was accepted with complete equanimity. It is true that somewhat later, when the prospect of British intervention became certain, the chancellor went through the motions of trying to hold Austria back, but he was half-hearted about it and at the last moment he cancelled the decisive instruction which might have caused Austria to recede.

(2) The stock German defense has always been that the possibility of mediation was spoiled by the general mobilization of the Russian army which, according to German argumentation, forced Germany to counter-mobilize and go to war. But the decision of the German government for war was taken before the news of the Russian general mobilization was received and before any answer had been received from Vienna as to whether Austria would negotiate. Furthermore, in spite of the alleged danger to Germany from the Russian mobilization, the German armies moved, not against Russia, but into France by way of Belgium, although they had no quarrel with France and no pretext for war except that France was the ally of Russia. Down to 1913 the German staff had alternative plans for attacking Russia or France, but it then decided that the only road to victory in a two-front war was to smash France first and then turn on Russia. Whether or not this was good strategy, it was certainly bad politics. The German general staff was the only one which insisted that mobilization necessarily meant war. Both Austria and Russia, the protagonists in the conflict, did not consider that mobilization precluded negotiation. It is on record also that the German minister of war was of the opinion that Germany could wait several days before reacting to the Russian general mobilization. The truth seems to be that the German staff had decided on war and used the Russian mobilization as an excuse.

Looking back forty years, the historian must note that the three fatal decisions of the July 1914 crisis—the German decision to support Austria, the Austrian decision to reject the conciliatory Serbian reply to its ultimatum, and the Russian decision for mobilization—were all sudden decisions, taken without reflection and damning the consequences. But as the Russian decision was the understandable consequence of the previous German and Austrian decisions, it is hardly fair to blame the Russians exclusively for their decision.

(3) But if the continental governments acted precipitately, the British government went to the other extreme and could not make up its mind what to do until war on the Continent was an accomplished fact and the German army had entered Belgium. Ever since 1914 the question has been hotly debated whether the war might

not have been prevented if Britain had promptly announced that it would stand by Russia and France. Neither historians nor diplomatists are agreed on the answer. I personally incline to the view that a British declaration at the very outset of the crisis might have saved the situation, but, since neither the British parliament nor the British public would have sanctioned such a declaration, the view remains academic.

In the years before the war, sometimes the hope was expressed that the Socialists of the various countries would heed the Marxist doctrine and refuse to obey the order of mobilization, and it is a fact that in July 1914 some German Socialists tried to get in touch with their French colleagues. But when the day came, every Socialist party accepted the official thesis that its country was the victim of unjustified attack and voted for the war credits. The German Social Democrats, to whom the Russian autocracy was anathema, asked no questions about the conduct of German policy and readily accepted the Russian mobilization as a sufficient *casus belli*.

Others expressed the hope that business interests would exert sufficient pressure on governments to stop them from war, and the record shows that in both Germany and England there was opposition to war in business circles. But their protests, which were not very loud, were ignored. Churches, too, were generally looked upon as forces for peace, but they also did not count in July 1914.

At the present time, the policy of the United States is based on the assumption that if we are sufficiently strong militarily, the Communists will not attack us. The same idea was constantly propounded before 1914. In fact, this was the invariable argument of all governments to persuade their peoples to pay the taxes necessary for maintaining great armies and navies. Yet the bloated armaments, far from creating a feeling of security, bred a feeling of insecurity. The greater the armaments, the greater the insecurity, and when the day came, no government hesitated from fear of possible defeat or from reluctance to face the horrors and misery of war. This is not to argue, as is sometimes done, that armaments caused the war. My point is that they did not prevent it. Remembering this, I do not feel so certain as do our present rulers that our super-bombs will forever intimidate our enemies.

So we come to the war itself. The expectation was widespread that the struggle would be brief. For one thing, it was often argued that modern war would be too expensive for governments to wage for a long time. This view derived from Norman Angell's famous book *The Great Illusion* published in 1910. Angell did not say, as is sometimes alleged, that economics would make war impossible. What he did say was that war would be ruinous to both victor and vanquished, and it was precisely for this reason that banking interests in July 1914 lifted their voices against the oncoming catastrophe.

More positive evidence of belief in a short war is the fact that no government had made serious economic preparation for war. In no country had reserves of food, fuel, or raw materials been accumulated. The German general staff believed that, if the neutrality of Holland were respected, Germany would be able to import freely through the Dutch ports. It was not until after the war had begun that the general staff addressed itself to the problem of organizing Germany's resources for war. Britain never had on hand food supplies for more than a few months, yet relied on the ability of the navy to control the seas and allow supplies to come in. In no belligerent country did the rationing of food begin until long after the opening of hostilities.

On the military side, it was the same story. The German general staff expected to defeat France in six or seven weeks—and they jolly well nearly succeeded!—and then turn on Russia, which was expected to collapse in a few months. This pre-war optimism of the Germans was matched on the Allied side after the war started, when there was much talk of the Russian "steam roller" getting to Berlin by Christmas. Lord Kitchener, the famous British soldier who was appointed secretary for war on the outbreak of hostilities, was not taken too seriously when he said that the war would last at least three years. Most interesting of all, no general staff had the least conception of the consumption of ammunition. By September, only a few weeks after the fighting had begun, the generals on both sides were complaining of the shortage of shells, although it had been assumed that there were supplies for months of campaigning. Apparently no government had planned in advance to make the enormous quantities of ammunition consumed by the war.

By the end of 1914, instead of the Germans getting to Paris or the Russians to Berlin, the war had reached a stalemate in Belgium and

France. The rival offensives of the French and German armies had both failed, and trench warfare had set in. This trench warfare lasted from November 1914 to March 1918. For more than three years millions of men on both sides lived in trenches fortified with concrete and defended by machine guns and mortars in the front line and by heavy artillery at the rear. The British and French generals, and to a considerable extent the German generals as well, managed to convince themselves that, if only enough shells were fired at the enemy trenches, these footholds could be demolished and then the infantry could go forward to seize the rubble that was left. The mirage of "break-through" fascinated equally Joffre and Haig and Falkenhayn, the principal French, British, and German commanders during the first two years of the war, and, in spite of repeated failures and enormous losses, they never admitted to themselves that the thing could not be done. In several instances cavalry was massed behind the lines to dash through into the open country. The inability of the professional soldiers to find a way out of this deadlock reflects rather sadly on their imagination. It is no wonder that a famous Frenchman, Georges Clemenceau, is reported to have said, after he became premier in 1917, that war was much too serious a business to be entrusted to generals.

If our Civil War was the first conflict in which railroads were used on a large scale, the First World War witnessed the *début* of the internal combustion engine. In land warfare, the gasoline engine was applied in three ways. First, from the beginning of operations motor trucks were used to transport troops, guns, ammunition, and supplies beyond the railheads, and sometimes, as at Verdun, this was the only means of transport to the immediate fighting zone.

Second, the airplane was used for the first time in war. At the beginning, it was used only for reconnaissance, for it was an untried weapon and there were only about five hundred planes in all the armies put together. Gradually it began to be used for locating trenches and "spotting" artillery fire. Then troops were bombed from low-flying planes, and ultimately high-flying planes were employed for general bombing. Not only this, but planes fought each other for the control of the air, which slowly but surely passed to the Allies, who were able to make more planes, thousands of them, than the Germans. This superiority did not, however, enable the Allies to overcome the advantage possessed by the Germans, who were, on the whole, content to remain on the defensive and let the Allies exhaust themselves in fruitless assaults against the trenches. These assaults did yield the Allies a few square miles of territory, but in March 1918, the line had changed little from November 1914, and the German armies were still firmly entrenched in Belgium and France.

The deadlock on the western front was in the end broken by the mechanical contraption which we call the tank. The conception of a self-propelling vehicle which could cross trenches and also shoot came from the brain of a subordinate British officer serving in France; this was the third application of the gasoline engine to the revolutionizing of warfare. But it was long resisted by the generals and it was carried through to execution largely by civilian initiative. Even after they had been converted to the idea and had begun to use the tanks in France, the generals refused to employ the tactics prescribed by the inventor, and it was not until November 1917 that the tank came into its own at the battle of Cambrai. Even then, the top German commander, the famous Ludendorff, was not impressed and failed to build German tanks in adequate quantities or to devise adequate defenses against it. Open warfare returned with the great German attack in the West in March 1918, and, thanks to superior numbers at the outset and to the use of liquid fire, they were able to break through. But when the Allies, now reinforced by the American armies, rallied, they possessed in the tank the weapon which enabled them to defeat the Germans. The Germans proved unable to improvise sufficient resistance, and being unable to defend their positions, they lost the war.

The war on the Eastern front was quite different from that in the West—"the unknown war," Sir Winston Churchill has called it. The territory fought over was vastly larger than the compact area of the West, and trench warfare never developed. The Russian army was large but after the first few weeks, short of munitions. The Austro-Hungarian army was a hodge-podge of rival nationalities, some of which were not loyal and deserted to the enemy in large numbers. Generally speaking, the Russians were able to defeat the Austrians, but were themselves beaten by the Germans. By the end of 1915 the Russians had been defeated and driven out of Poland and, although they made a brief comeback in 1916,

they had suffered so greatly that by 1917 they were ripe for revolution.

Yet we must not underestimate the part played by the Russian army in the war. In the opening campaign, the Russians chivalrously invaded East Prussia, in order to help their French allies, and, although they were badly defeated, they so scared the German general staff that it detached two corps from the German armies in the West and sent them eastwards. Had these troops been available for the battle of the Marne, which stopped the German army before Paris, the result might have been different. Later, even after the Russians had been defeated, the German general staff never dared denude the Russian front and throw all its forces into the battle in France. I have often wondered what would have happened if in 1914 Germany had stood on the defensive in the West and launched its big offensive against Russia. Probably the German army could have stopped the French offensive, and certainly, if Belgium had been left alone, Britain would not have entered the war at the start.

The most spectacular incident of the war was the campaign at the Dardanelles. This arose out of a Russian request to have the Turkish pressure in the Caucasus relieved, which came at the moment when the Allied leaders were seeking a way out of the trench warfare in the West. Strategically, the concept was sound, for success would not only have put Turkey out of the war but would have opened up a route to Russia far easier than the difficult voyage to Archangel. It would take too long to list the numerous mistakes by which the expedition was bungled from the start. In spite of these errors, success was narrowly missed. After the great naval attack of March 18, 1915, the Turkish ammunition was exhausted and plans were made to evacuate the peninsula and Constantinople when the attack was renewed the next day; but the Allied admiral had been scared by his losses and failed to go on. During the land fighting, the British almost captured the critical heights which dominated Gallipoli; but bad generalship and bad luck were too much for them. It is generally agreed that if the Allies had been able to open the Straits and supply Russia with what it needed, the Russian revolution might not have happened.

The Russian revolution of March 1917 had a double aspect. On the one hand, it was a revolt against the corruption and inefficiency of the tsarist regime, a revolt brewing for half a century and already once attempted, in 1905; on the other hand, it was a protest against continuing the war. As a movement to achieve liberty and reform, it was hailed by the Allies—"heartening news" Woodrow Wilson called it—who had always been somewhat uncomfortable about the Russian autocracy and hoped that upheaval would find echoes among the discontented peoples of the Central Powers, as indeed it did. The new order forced the Austrian government to summon the Austrian parliament, which had been in eclipse since 1914, and in Prussia loud demands began to be heard for reform of the antiquated three-class franchise. The new Russia was, to quote Wilson again, "a fit partner for a League of Honor."

The revolution as a protest against continuing the war was not understood in the West. The British and French were being hard pressed on the Western front and they refused to face the possibility of Russia's withdrawal from the war; they even believed that the revolution would strengthen Russian enthusiasm for the war. The demands of the Russian radicals for peace on the basis of no annexations and no indemnities were deeply resented, and when the Provisional Government set up in March, which was keenly aware of the groundswell calling for peace, asked for a reconsideration of war aims, no reply was returned. Under pressure from the Allies, the provisional government staged an offensive in the summer of 1917, which, though momentarily successful, was soon beaten back, and the failure not only completed the demoralization of the Russian army but also deprived the government of reliable troops with which to resist the Bolshevist attack of November 7. Whether greater understanding by the Allied governments of the Russian situation would have prevented the Bolshevist triumph, no one can say, but their actual policy played directly into the hands of the Communists. On the other hand, it is only fair to remember that if the Allies had agreed to discuss peace on the Russian terms, the Germans might have felt it safe to shift troops from the Eastern to the Western front, whereas actually they did not begin to shift troops until much later, until they had opened negotiations with the Bolshevist government.

The German general staff quickly realized that the first revolution had profoundly affected the Russian army; so it refrained from attacks and sent it own agitators, including Lenin himself, to stir up agitation among the troops. Undoubtedly

the German maneuvers contributed considerably to the breakdown of the Russian army, and Germany profited at first from this. Not only was it safe to begin transferring troops to the West, but, at the end of 1917, Germany was able to open negotiations with the Bolshevist government which resulted in the Peace of Brest-Litovsk. By this treaty Russia was forced to surrender vast areas of its territory to the Central Powers and to accept harsh economic terms. The triumph of Germany seemed complete. Yet, in the end, the Germans were hoist with their own petard, for, after the establishment of diplomatic relations between Germany and the Communist government, the Russians were able to launch their Communist propaganda in Germany and this helped undermine the German will to resist in the autumn of 1918. So the Allies had no monopoly on not knowing how to deal properly with the situation created by the Russian revolution.

What were the war aims of the Allies which the Russians wished to discuss? Americans will immediately think of the secret treaties of 1915–1917, by which the Allies planned to cut back the frontiers of Germany and distributed the overseas possessions of Germany among themselves, promised parts of Austria-Hungary to Italy and Romania, awarded Constantinople to Russia and partitioned most of the Ottoman Empire. These agreements, secretly negotiated and carefully not published, were certainly not in harmony with the professions of disinterestedness and respect for the principle of self-determination which the Allies publicly proclaimed. But, in fact, they were in large measure the consequences of military situations. The promise of Constantinople to Russia was made when the Allies were trying to force the Dardanelles; if they had succeeded and captured the city, they would have aroused deep suspicion in Russia had they tried to hold it for themselves. Was it unreasonable to partition the Ottoman Empire? The Allies had offered to guarantee its integrity in return for Turkish neutrality; but Turkey had joined the enemy and vastly extended the scope of the war. The promises made to Italy and Romania were exaggerated and unjustified, but there again it seemed that the tide of battle might be turned if these new armies entered the conflict. To some extent the Allies were in the position of parents who had to ransom kidnapped children—they did not haggle too much over what they had to pay. The secret treaties contained many unjustified provisions, provisions which, when the time came to put them into effect, proved impracticable. Actually, the treaties were of less consequence than is often alleged, and since they were published by the Bolshevist government of Russia at the end of 1917, they had to face the test of public opinion.

If the Allies in 1917 lost one ally, Russia, they gained another, the United States. The outbreak of war in 1914 took the United States by surprise, for American opinion had little notion of the political problems of Europe and accepted at face value the surface indications of a stable and peaceful Europe. For more than two years, American opinion favored neutrality, and Woodrow Wilson was reelected President in 1916 because he had kept the country out of war. Nevertheless, the great majority of Americans had early in the struggle decided that Germany was primarily responsible for the war, and when the German army marched into Belgium and committed atrocities (which were no doubt exaggerated), Germany appeared to be the Beast of the Book of Revelation. This bitter hostility to Germany was perhaps the decisive factor in American policy.

As soon as the war started, the United States was affected by the British blockade of Germany, and early in 1915 it was affected by the submarine warfare inaugurated by Germany. Both sets of measures were regarded by the United States as contrary to international law, and our government constantly protested against both. But the protests against the British blockade never had teeth in them, whereas in the case of submarine warfare the United States announced that it would hold Germany to "a strict accountability." In other words, the United States government set up two standards of conduct, a much sterner one for Germany than for Britain. We took the view that goods stopped or seized by the British could be paid for, but human lives lost by submarine action were irreparable. This discrimination was approved by the vast majority of the American people. In passing, it may be noted that the submarine was the fourth way in which the gasoline engine was applied to the revolutionizing of warfare, the others being, as already noted, the truck, the airplane, and the tank. The Germans argued that the use of the submarine was a legitimate reprisal for the British blockade, but the United States government refused to admit a connection between the two illegalities.

Wilson hoped that his threat to break off relations with the German Empire would persuade the German government to give up the submarine warfare, and the German civil authorities were willing; but they were overborne by the military, who insisted that by no other means could Germany win the war. Because of the controversy aroused by the submarine, Wilson was reluctant to make its use a *casus belli* for the United States, and he did not go to war until American lives had been lost in American ships sunk by submarines. In Wilson's eyes, the real issue was much larger. He was persuaded that the peace of the world could be safeguarded in the future only by the organization of a League of Nations and that the military and despotic regime of Hohenzollern Germany had to be destroyed before a sound peace could be made or a League of Nations organized. A more immediate and practical consideration was that the United States, in its own interest, could not allow Britain and France, who had been unable to defeat Germany, to go under.

During the 'thirties it became fashionable, in some quarters, to say that the United States had been tricked into the war by Allied, more particularly British, propaganda. I do not believe it. Certainly it was not true of Wilson. The record leaves no doubt that at the end of 1916 he was greatly annoyed with the British, partly because of the British blacklist of American firms trading with Germany and the interference with the mails, partly because they had not taken kindly to his peace move of December 18. The record also shows that if Germany had not double-crossed him with its peace maneuvers, if Germany had met him half way in the matter of peace, Wilson was prepared to exert pressure on Britain to discuss peace with Germany; all he had to do was to stop American credits for British war purchases, for the British were about at the end of their rope. Nor were the American people made anti-German by Allied propaganda. That was done by the German government and the German army. What British propaganda did do, I suggest, was to dispel the anti-British feeling more or less latent in the United States ever since the Revolution.

Another version that enjoyed some currency during the years of the Great Depression was that Big Business, more especially the munitions makers, took us into the war. It is a fact that when the war broke out in 1914, the United States was in the grip of a business depression, and that we owed a lot of money to Europe. By the end of 1916 the country was revelling in prosperity, thanks to the enormous purchases of food and war supplies by the Allies; not only had we paid off our debts, but Europe had become our debtor. Therefore, it was argued, in order to insure the payment of these debts and to keep up the flood of war orders, Wall Street maneuvered us into the war. I have never seen the slightest evidence to substantiate this view, which I never heard expressed in 1916–1917. Such an argument would, I am sure, have had no appeal for Woodrow Wilson. It is, indeed, ridiculous to contend that the United States went to war for the sake of economic gain. Quite apart from the fact that it would have been cheaper for us to keep out of the war, persons whose memories go back to 1917 will, I am certain, agree that we were animated largely by idealistic motives. Maybe we were not so unselfish as Wilson asserted, for it was certainly to our national interest that Germany should be defeated, but we were not thinking in terms of dollars and cents. I recall that when the charge of Wall Street manipulation was aired, Thomas W. Lamont, of the firm of J. P. Morgan & Company, wrote an indignant letter to the *New York Times* repudiating the very suggestion. He pointed out that he had sons of military age and it was preposterous and inhuman to suggest that he would be willing to risk their lives for the sake of gain that might allegedly come to him from American participation in the war. It is possible to argue that the United States made a mistake in going to war (although I do not think so), but not that we did so for sordid motives.

The war was transformed by the entry of the United States from a conflict between two rival groups of acquisitive powers into a struggle between the democratic and peace-loving peoples of the world and the reactionary, military regime of Germany. Wilson's war address of April 2, 1917, was implicit with the idea of self-determination, the right of peoples to determine their own political allegiance and form of government. When the Bolshevist government of Russia at the end of 1917 published the secret treaties already referred to, Wilson particularized his program in the famous Fourteen Points, a program which, whatever may be thought of its details, made Wilson's name one to conjure with throughout

the world. Although the Fourteen Points were not at that moment formally accepted by the Allied governments, our Associates in the war, their appeal to the peoples of the Central Powers and to neutrals was very great and obtained for him a position of moral authority never before or since enjoyed by any statesman. It is well to remember that, in addition to Wilsonian idealism and German militarism, a third ideology was competing for the favor of mankind—the call of Communist Russia. The Bolshevist leaders of 1917 were persuaded that the masses throughout Europe and especially in Germany were only waiting for the chance to overthrow their governments and usher in the Communist millennium, and they did their best to excite the workers to revolution. But their propaganda evidently made small appeal, for when the collapse of 1918 came, there was little enthusiasm for Communism, and what little there was was easily beaten down. For this the preachments of Woodrow Wilson had no small part of the credit.

The American contribution to the military winning of the war was very great. In France the tired British and French armies almost broke under the heavy German attacks in March 1918; it was the fresh, huge American army that provided the superiority necessary for victory. No doubt our raw troops made many mistakes which the seasoned British and French had learned to avoid, but once our troops went into action on a grand scale, the Germans, although they continued to fight well, knew that the war was lost. At sea we introduced new methods which proved successful in beating the submarine. While it would be untrue to claim that we won the war, clearly the war could not have been won without us.

American policy contributed enormously to the ending of the war in 1918. From the moment that the United States entered the war, the discontented nationalities of Austria-Hungary had looked to Wilson as the prophet of self-determination, and they were disappointed when in the Ninth of the Fourteen Points, he talked only of autonomy within the Habsburg state. At that moment, it seemed possible that the Emperor Charles might be willing to negotiate for a separate peace, so Wilson had to be cautious. But the negotiations came to naught, and presently the United States recognized publicly the claims for independence and unity of the Czechoslovaks, Jugoslavs, and Poles, the three most important minorities of Austria-Hungary. This was the beginning of the end. At the beginning of September, the Austro-Hungarian government, now *in extremis*, appealed to Wilson for the opening of peace negotiations on the basis of the Ninth Point, only to be told that the Ninth Point had been overtaken by events. In other words, the United States no longer stood for autonomy or federalization within Austria-Hungary but sanctioned its destruction. This was the signal for which the nationalities had been waiting, and when the Austrian army was defeated by the Italians at Vittorio Veneto, first the Czechs and then the Jugoslavs, and finally even the Magyars declared their independence. When the armistice with Italy was signed on November 3, the Austro-Hungarian state had ceased to exist. The Czechs honored their liberator by naming the principal railway station in Praha for Wilson.

On October 4, after the Allied armies had broken the Hindenburg Line in France, the German government appealed to Wilson for an armistice, this to be the preliminary to a peace based on the Fourteen Points. Just previous to this the Kaiser, by a stroke of the pen, had taken the revolutionary step of making the imperial chancellor responsible to the Reichstag. The proposal for an armistice emanated from the general staff, not from the civil government, which realized that peace on the basis of the Fourteen Points would signify defeat for Germany, and the new chancellor, Prince Max of Baden, opposed the move. Ludendorff, however, was adamant, and the appeal was made. But Ludendorff was not sincere. He was playing for time, in hope of reconstituting the German armies and fighting on. Wilson sensed this and acted accordingly. In a series of notes, about which he did not consult the Allies, he demanded that the German armies evacuate all Allied territory and stop the submarine warfare; he also challenged the authority of the German government and declared that the Allies could negotiate only with a democratic regime. Finally he stated that, if he had to negotiate with the military rulers and the monarchical autocrats of Germany, he must insist not on negotiations but on surrender. Pushed from one position to another and realizing the hopelessness of a *levée en masse* as urged by Ludendorff, the German government accepted all of Wilson's conditions. This masterly diplomatic

achievement, done publicly, was accomplished in the face of considerable criticism at home and in the Allied countries, where it was feared that Wilson would let the Germans off too easily. All things considered, this must be reckoned Wilson's most brilliant performance.

American diplomacy was equally effective in dealing with the Allies. When Wilson communicated to them his correspondence with the German government, they were reluctant to accept the Fourteen Points, but Wilson's representative in Paris, Colonel House, threatened that, in case of refusal, the United States might negotiate a separate peace with Germany. Whether House was authorized to say this or did it on his own responsibility is not clear; but the effect was magical: the Allies accepted the Fourteen Points, with two reservations. The American commander, General Pershing, was opposed to an armistice, but the British and French commanders favored it, and Marshal Foch, the commander-in-chief, asserted that if the Germans accepted his conditions, they would be rendered powerless. Wilson, for the only time, did not back up Pershing and followed Foch. The Germans did accept the Allied terms, and the war came to an end on November 11, 1918.

Or rather it came almost to an end. During 1918 Allied troops had landed on Russian soil at Archangel, in the Caucasus and in Siberia. The alleged object of these expeditions was to protect Allied interests against Germany, but the Bolshevist government regarded them as directed against itself and it broke off relations with the Allies. In October, however, as the war was nearing its end, the Bolshevist government addressed itself to Wilson and proposed negotiations. No answer was sent to this communication. But the armistice did require Germany to denounce the Treaty of Brest-Litovsk and to withdraw its troops from Russia. The Allies hoped that the Bolshevist regime would be overthrown (for civil war was already raging in Russia) and they were trying to safeguard their relations with a future "white" regime. Thus at the end of the war, the position of Russia remained uncertain.

The war of 1914–1918 was the bloodiest and most costly yet fought in modern times. Exact figures of the casualties are not obtainable and, on so vast a scale, do not matter. It is safe to say that about ten million were killed, something like twenty million were wounded, and another six or seven million were listed as prisoners or missing. The money cost has been estimated as in excess of two hundred billion dollars. In addition, vast areas in Belgium and northern France, Poland, Russia, and the Balkans were devastated and millions of tons of shipping were lost at sea. The people of all the warring countries in Europe suffered hardships on a scale hitherto unknown. Until the greater losses and suffering of the Second World War had to be endured, it was believed that human patience had been pushed to the limit. In the minds of governments and people alike the losses and the sufferings were justified by the belief, or at least the hope, that they were the necessary preliminaries to the New and Better World that was to emerge from the war. When, later, it became doubtful whether a better world had emerged, many asked the question whether it would not have been wiser, instead of fighting the war to a finish, to have negotiated a compromise peace. That was Wilson's original idea, the explanation of his phrase "peace without victory." It was only after Germany had double-crossed him at the beginning of 1917 that he swung over to the view that Germany had to be beaten down. When the German government proposed peace negotiations at the end of 1916, Hohenzollern, Habsburg, and Romanov still sat upon their thrones and the framework of European society had not been shattered; Communism had not yet come to power in Russia. But the German terms, as made known secretly to Wilson, were not the terms of compromise. Nor were the Allied terms, as publicly stated, the terms of compromise. What Wilson might have accomplished, had he been able to play his hand, no one can say. So it came about that the war was fought to a finish. The Central Powers were overwhelmed, and, when the Allied governments assembled at Paris in January 1919, to make peace and to create the New World, they had pretty much a free hand.

The world in which we live derives directly from the First World War. In the first place, the peace treaties concluded after the war redrew the map of Europe west of Russia in revolutionary style. Whether this was wisely done has been disputed ever since; nevertheless, except in a few details, it survived the Second World War. Even in Eastern Europe, most of the states set up after 1919 still exist, even though their boundaries have been much changed by the Soviet Union

since 1945. Secondly, the First World War destroyed the balance of power as it had existed in Europe since the Congress of Vienna, and neither in the twenty years between the First and Second World Wars nor after the latter struggle has a real balance been set up. In the twenties and thirties no balance was possible because Soviet Russia had in large measure retired from European affairs; no balance is possible today because Soviet Russia is now more powerful than the rest of Europe combined. Thirdly, the First World War destroyed the tsarist régime in Russia and paved the way for the establishment of Communism, which today presents the greatest problem to Europe and the whole Western world. It is for these reasons that I have ventured today to recall to you some of the most important aspects of the war of 1914–1918 which, up to that time, was the most terrible conflict that the modern world had witnessed.

A STRATIGRAPHIC EXCAVATION AT EL RISCO, MEXICO

WILLIAM J. MAYER-OAKES

Assistant Professor of Anthropology, University of Toronto

CONTENTS

PAGE

Introduction ... 332
Site description ... 333
Excavation ... 333
Artifact description and analysis ... 336
Background of study ... 336
Procedures ... 337
Pottery types ... 339
Form analysis ... 345
Rim analysis ... 346
Special data ... 347
Figurines ... 350
Non-ceramic data ... 352
Periodization ... 352
Burial description ... 355
Comparisons ... 356
Cerro Portesuelo ... 356
Tolstoy's survey ... 364
Site interpretation ... 365
Social structure hypothesis ... 365
Salt-making theory ... 367
Population inferences ... 368
Settlement pattern inferences ... 368
Comparisons ... 370
Future research ... 371
Acknowledgments ... 372
References ... 372
Appendix 1 ... 373

INTRODUCTION

In February, 1951, my wife and I briefly visited various areas of central and southern Mexico for the purpose of obtaining a photographic record of prominent archaeological sites. While driving on the Pan-American highway just north of the Mexico City limits we noticed that the road cuts through several low mounds filled with bright red sherd refuse. Examination of this road cut material and considerable more collected from the surface of the extensive site traversed by the highway demonstrated the presence of an amazingly high proportion of an unusual Mexican pottery type. Briefly, the pottery was uniformly red in color, poorly fired, fibre-tempered, and covered on its exterior surfaces with fabric-impressions. Associated surface sherd material was largely utilitarian and non-distinctive, but typical Aztec Black/Orange sherds, sherds suggestive of Teotihuacan and Archaic cultures, as well as Archaic figurines were collected in very small numbers.

Thinking that we had found something quite unusual which offered a possible link with North American prehistoric cultures, we resolved to return for intensive work at some future date. Two months later James B. Griffin pointed out to me the fact that this unusual pottery had, in fact, been recognized for some time, the most recent publication concerning it being Griffin & Kriegers' 1947 article in *American Antiquity*. Properly chastened about the "newness" of our discovery we were nonetheless struck by the lack of knowledge about this particular unique type of pottery and, what has since become the paramount problem, the general inadequacy of knowledge about ceramic sequences in the Valley of Mexico subsequent to the Archaic or Formative period.

Discussion and correspondence with various Middle American specialists demonstrated the fact that the fabric-marked pottery was generally considered late in time, but its exact chronological and cultural affiliations were not clear. Several unexplained occurrences in definitely early contexts were disturbing possibilities for North American contacts.

We decided to return to Mexico for excavation and survey work aimed at resolving this problem. Plans and arrangements for a Carnegie Museum expedition were made in 1953 and we arrived in Mexico City on January 1, 1954. Surface survey work to determine distribution of the fabric-marked pottery was started on January 5 and continued for three weeks. From Gordon Ekholm of the American Museum of Natural History we had learned that Paul Tolstoy, a Columbia University graduate student, was in Mexico doing a seriation study of the northern Valley of Mexico. Contacting Tolstoy early in January we were able to shorten our survey period and concentrate on excavation. His project, of greater duration than ours, would include

the survey phase of our program. It would be advantageous for us to spend most of our energies and time on an excavation which, then, could be interpreted in the light of the broad surface survey conducted by Tolstoy and upon which a seriation could be based. Since he planned no excavations, our two independent projects might profitably be correlated and would perhaps supplement each other. We agreed to cooperate in this fashion and our two-man party was happily augmented for most of the excavation work by Paul Tolstoy and his wife, Christie.

Our own survey work resulted in the recording of ten major sites producing the fabric-impressed ware. By combining the knowledge gained from these sites with the information offered by Tolstoy, we were able to decide on an excavation site. And, strangely enough, we came back to the original find spot, the site on the Pan-American highway, El Risco. Actually, we selected a single mound or tlatel at the south end of the site and to distinguish it from the rest of the large occupied area, referred to it as "El Risco II."

SITE DESCRIPTION

Map 1 is an artist's conception of the Valley of Mexico, emphasizing the topographic complexity of this high basin and range area. The pre-Conquest lakes, however, are shown as by Vaillant (1944) in the front and back end paper maps. Various well-known and pertinent sites are indicated, symbolically, while the numbered dots represent sites producing the fabric-impressed pottery. El Risco II can be seen as dot no. 6 on the southwestern shore of former Lake Texcoco.

Map 2 is based on the Texcoco sheet 14 Q-h(2) published in 1952 by the Comision Cartografica Militar. It shows in topographic detail the area surrounding the El Risco site. The El Risco II location can be seen as a dot at the end of the most southeastern spur of Cerro Chiquihuite. The entire El Risco site is not indicated on map 2 but it extends from station II nearly to the village of Xalostoc, and can be characterized as a series of low mounds or tlateles on the edge of or slightly within the boundaries of the former lake.

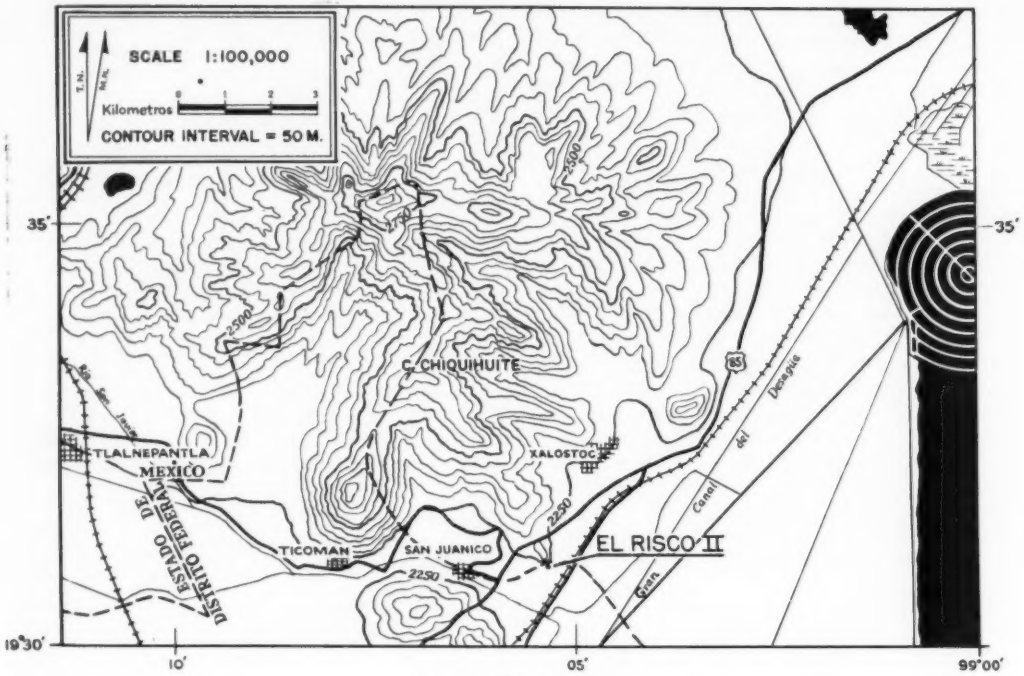
A schematic aerial view of El Risco II is shown in figure 1. Our excavation, Cut A, was made in the central peninsula, all that remained of a man-made extension to the hill (known locally as El Risco) shown rising to the west. This had

MAP 1. Valley of Mexico.

evidently been a tlatel or chinampa-type extension of land surface as described by Apenes (1943). Excavation for a railroad cut (not shown on map 2) had carried away most of the western part of this mound. Additional excavations, perhaps for road fill, had removed the eastern portion of the site. The central portion left standing was highest at the northern end and tapered down to the former lake bed level at the south.

EXCAVATION

On January 25, 1954, we began our stratigraphic cut at El Risco II by selecting a portion of the eastern face of the peninsula for a profile. This was cleaned and made vertical in order to give us an open side to our cut as well as a preliminary view of what the excavation would produce. An area 10 feet wide by 12 feet deep was cleaned, examined, photographed, and sketched as the east wall profile. This profile is shown in fig. 2.



MAP 2. El Risco area.

The surface material here was predominantly fabric-impressed pottery and this ware was noted in quantities throughout the upper half of the deposit as seen in the east wall. The lower half was equally rich in sherds, but apparently not the fabric-marked style. Because this spot appeared to contain at least the early stages of the fabric-impressed pottery, we decided to continue our work here. A rectangle 10 feet long and extending 6 feet back into the peninsula was laid out and, since we knew in advance that at least a 12-foot depth was assured, excavations were commenced in arbitrary one-foot levels, measuring from the approximately level surface.

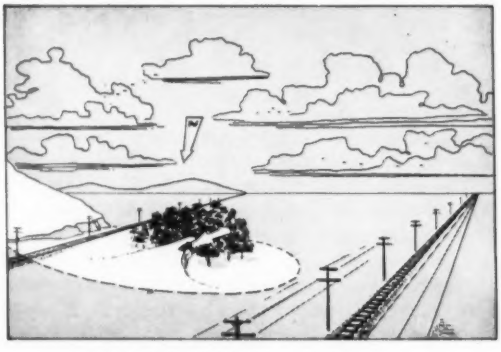


FIG. 1. El Risco excavations.

In our approach to the techniques selected for this work we were guided by what we thought to be inadequacies in previous Valley of Mexico investigations. Two primary factors were considered. First, many excavations have been done by local workmen under the supervision, from time to time, of trained archaeologists. Using this technique the chances for errors of both omission and commission are fairly good. We lessened these chances by doing all the actual labor of excavation ourselves. Second, only rarely is screening used as the method for selecting items to be saved from the excavation. Samples selected only by the method of hand searching are apt to be biased in favor of larger and easily recognizable items. We approached the solution of this problem by the use of a ½-inch mesh wire screen through which all soil from the excavation was sifted.

In defense of the techniques of using unskilled

workmen and no screening, we point to the large numbers of specimens that can be obtained by these methods. Our slower, more meticulous technique produces so many more specimens, however, that it appears to be feasible, from the study and analysis point of view, only for small-scale work in the Valley of Mexico.

Our method, thus, was designed to derive for us an exhaustive and accurate sample from a very limited area which had been chosen to help solve a particular set of problems.

A third factor considered in our selection of site had to do with the type of occupation represented. Most excavations in the Valley of Mexico have been conducted at important urban or religious centers. We wanted to find out something about the nature of a simple rural or suburban archaeological deposit and at El Risco we had what was evidently such a rural community. This was the "set of problems" mentioned above as a determinant of our method.

From the profile drawing (fig. 2) it is apparent that the deposit at Cut A was not uniform from top to bottom. The sloping deposits above the 5-foot depth were quite apparent, while the horizontal deposits below this depth are equally apparent. The upper of the two horizontal bands of "lighter-colored, coarse-grained soil" seems to mark an unconformity in the deposit. Correlations of this physical stratigraphy with cultural changes will be discussed below. Suffice it now to suggest that in the late stages of occupation here, the central section of the tlatel was to the north and west with the upper five feet of our cut being on the sharply dipping beds of the deposit edge. Below the 5-foot level our cut appears to be in the central part of the deposit where materials were laid down in horizontal strata.

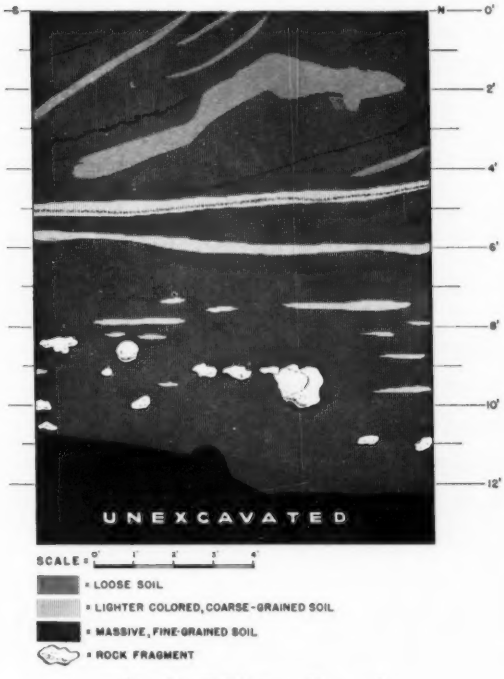


FIG. 2. El Risco profile, cut A.

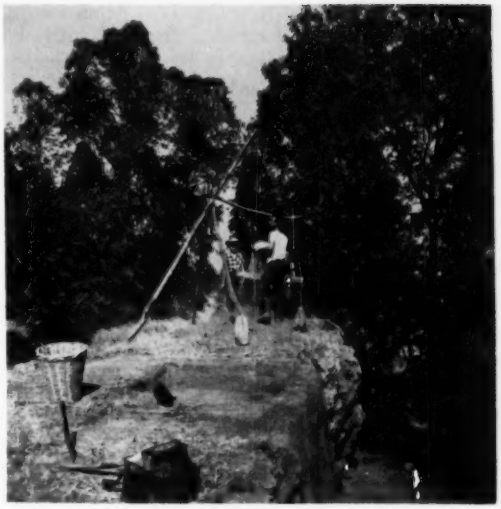


FIG. 3. El Risco, cut A, upper levels.

Noting this sloping character of the upper part of the deposit we split each of the first five levels into a north and south unit, keeping the materials separate in the hope of demonstrating internal time changes.

Because this excavation was made in the winter the soil was very dry and, below the top two feet, quite compacted and hard. Dust was a problem, but because of the open nature of the site, we did not require special respiratory equipment although there were occasions when we could have used it. For the first three levels digging technique consisted of shovel cuts to fill large baskets for carrying to the screen which had been set up on top of the tlatel (fig. 3). After the third level it was possible to set the screen up on the surrounding floor below the cut

and shovel directly into it (fig. 4). At this time, however, it became necessary to use handpicks to loosen the soil before it could be shoveled. This method was the only one feasible in the tough, compacted soil. In addition, the pick method was found to be less damaging to sherds which seemed to make up about 50 to 70 per cent of the volume of the cut in most levels.

Major features encountered were the two horizontal bands which were generally more compact than the rest of the deposit. They were not "peeled" off, but appeared to be relatively sterile. Large rocks were scattered between the 8–11-foot levels. These were especially concentrated in the central part of the cut between 9 and 10 feet where they were associated with and covered a flexed burial, described below.

The base of our excavation was determined by an irregular travertine-like layer of caliche which covered the bedrock underlying the site.

Excavation of the levels in the cut started on January 29 and was completed on February 13; eleven days were spent on the dig with an average crew of three. Five additional days were spent washing the finds and, before returning to Pittsburgh with the entire collection, we spent 10 days in preliminary sorting of the top, bottom, and middle levels.

FIG. 4. El Risco, cut A, lower levels.

ARTIFACT DESCRIPTION AND ANALYSIS

BACKGROUND OF STUDY

A preliminary sorting of selected levels from the site was done in Mexico prior to shipping the collection to Carnegie Museum. The purpose here was primarily to provide a preliminary view of the stratigraphic range and was more useful at that time to Tolstoy than to me. The collection of 750 pounds of pottery arrived in Pittsburgh early in April, 1954. Definitive sorting was started soon thereafter by my wife and continued into May when the work was suspended owing to pressure of local field activities. By the fall of 1954, when we returned to the museum, family responsibilities and final writing of the Upper Ohio Valley survey report combined to reduce descriptive study and analysis of Mexican materials to a bare, desultory minimum.

By May of 1955 the upper seven levels had been sorted and typed. I presented a progress report on this work at the Society for American Archaeology annual meeting held at Bloomington, Indiana, in May, 1955. Local field activities once again forced suspension of El Risco studies. It was not until April, 1956, that we could resume our work in earnest. By then Tolstoy's study had been completed and accepted as a Columbia University doctoral dissertation.

In the period from April to August, 1956, the final studies of sherd material were accomplished. All written data and rough manuscripts were moved to the University of Toronto in September of 1956. Final writing took place there in the winters of 1956–1957 and 1957–1958.

Although there are some discrepancies and differences, the type descriptions derived by Tolstoy are comparable to our categories. We will thus not repeat the physical descriptions (in which we agree for about 99 per cent of the cases) he has prepared for publication. Instead we will use these same types, describing only the physical differences apparent from our excavated material and the modifications in interpretation based on our somewhat different analysis.

The primary problem in using the types defined by Tolstoy centers on his use of different type names for very similar types in the adjoining Classic and Post-Classic periods. He has made sharp time distinction on the basis of his surface-collected materials and feels that the form changes, particularly, are of primary importance in supporting this distinction.

Because the upper levels at El Risco produce well-known Aztec materials, there is no question, for instance, that our "Brown" type in these

levels is Tolstoy's "Texcoco Brown" of the Post-Classic period and not "San Martin Brown" of the Classic period. There is also little question of the cultural continuity in our cut from top to bottom. The apparent depositional interruption noticed in the profile at the 5–6-foot level is not reflected in significant type displacements or changes. As will be shown below, however, there is definitely cause to reconsider some of the Classic and Post-Classic type distinctions in the light of the changes which occur to some of these Post-Classic types during the time span of Cut A at El Risco II.

PROCEDURES

Sorting was carried out utilizing surface finish texture to separate the fabric marked from all other wares. Burnished surface or slip color was used to distinguish most of the remainder, but plain sherds were distinguished by the lack of slip or burnishing, while all decorated sherds were separated out whether decoration was by painting, incising, stamping, or other means. Figure 5 shows the general distribution of these major categories through the entire depth of the excavation.

The color sorting of sherds was done on a rather broad general designation basis, broken down by categories which had obvious internal distinctions (e.g., "Orange" had two definite sub-groups; "Red/Buff" had two definite sub-groups).

The initial sorting was done by levels, systematically, from top to bottom. This phase of the study was accomplished during the intermittent study period from April, 1954, to April, 1956. The bottom three levels were sorted after Tolstoy's study had been completed and he had graciously sent me a copy of the manuscript. I was, thus, able to compare in the final stages of my sorting process with his final types. We were in almost complete agreement. The only differences were: (1) Tolstoy had eliminated an earlier "Grey" category to absorb it in "black" and "brown." We retained it. (2) We found no "filleted" category. (3) We did find in minute amounts several decorated varieties not listed by Tolstoy.

After sorting for general color and surface finish categories, I went back over all the speci-

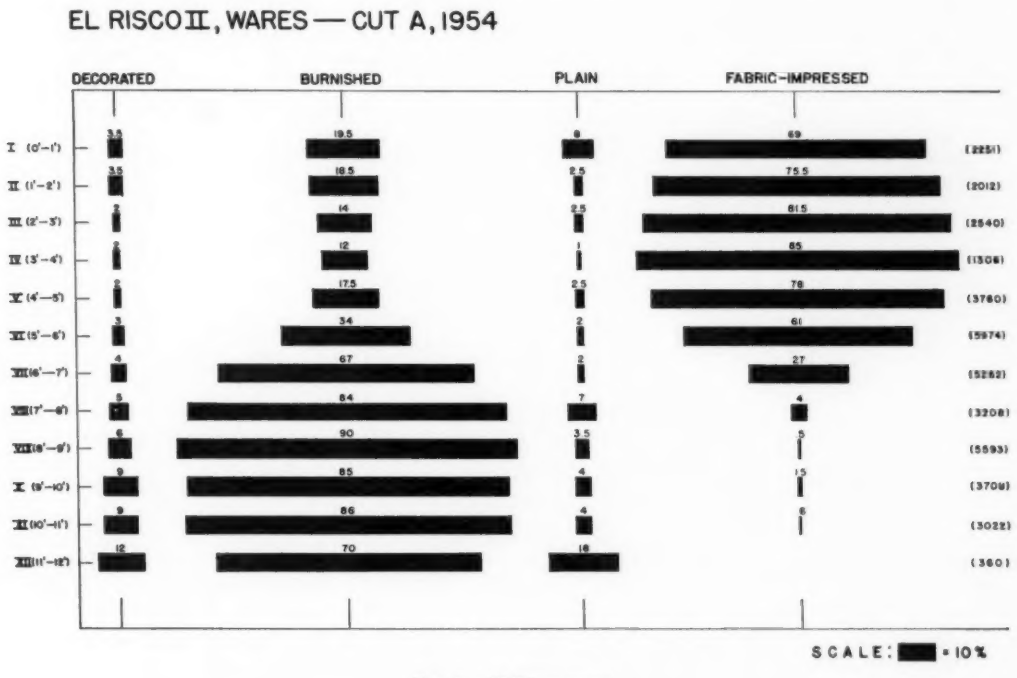


FIG. 5. El Risco wares.

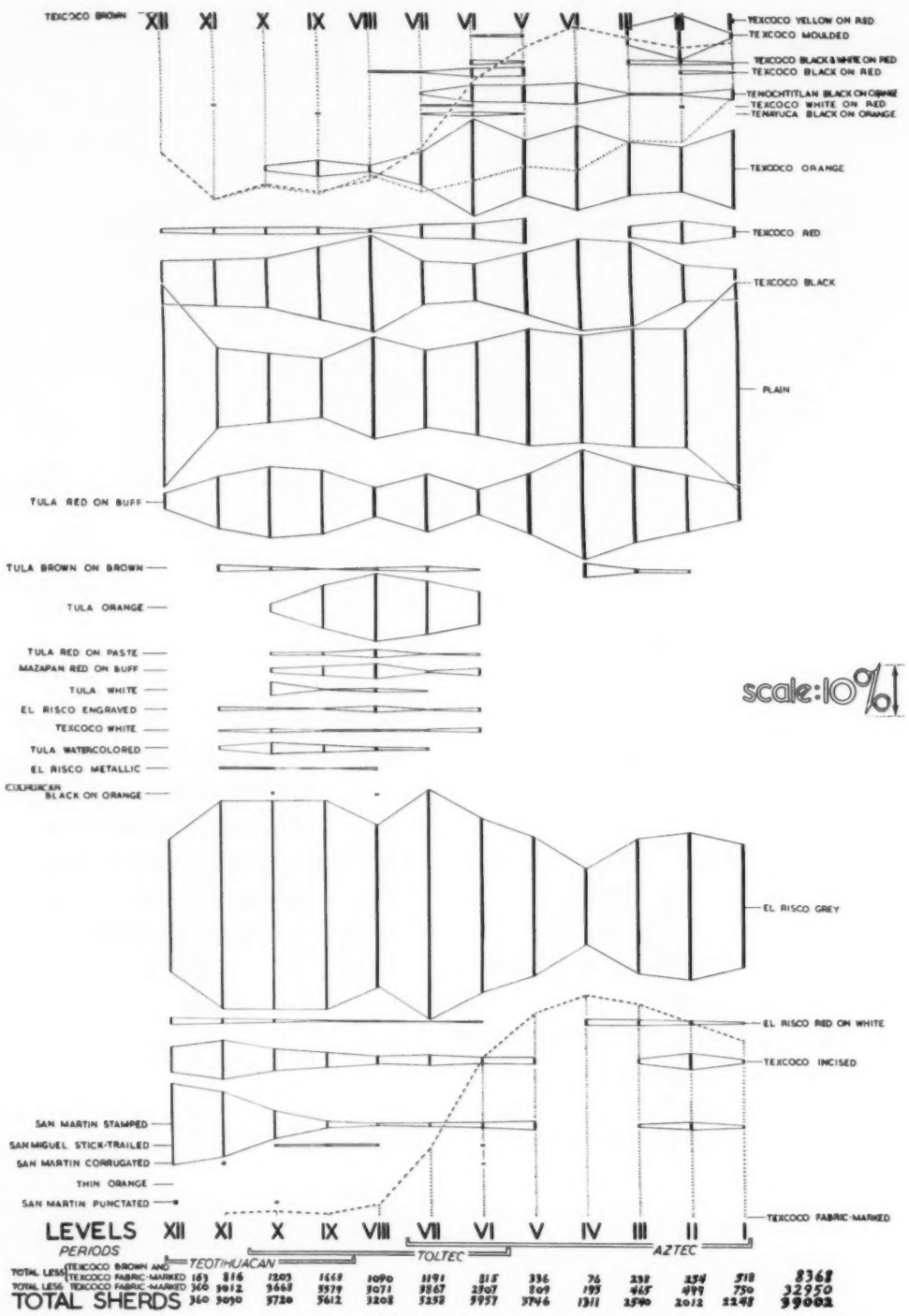

TEXCOCO BROWN
XII XI X IX VIII VII VI V VI III II I
TEXCOCO YELLOW ON RED
TEXCOCO MOULDED
TEXCOCO BLACK & WHITE ON RED
TEXCOCO BLACK ON RED
TENOCHTITLAN BLACK ON ORANGE
TEXCOCO WHITE ON RED
TEMAYUCA BLACK ON ORANGE
TEXCOCO ORANGE
TEXCOCO RED
TEXCOCO BLACK
PLAIN
TULA RED ON BUFF
TULA BROWN ON BROWN
TULA ORANGE
TULA RED ON PASTE
MAZAPAN RED ON BUFF
TULA WHITE
EL RISCO ENGRAVED
TEXCOCO WHITE
TULA WATERCOLORED
EL RISCO METALLIC
CULHUACAN BLACK ON ORANGE
scale: 10%
EL RISCO GREY
EL RISCO RED ON WHITE
TEXCOCO INCISED
SAN MARTIN STAMPED
SAN MIGUEL STICK-TRAILED
SAN MARTIN CORRUGATED
THIN ORANGE
SAN MARTIN PUNCTATED
TEXCOCO FABRIC-MARKED
LEVELS XII XI X IX VIII VII VI V IV III II I
PERIODS
TEOTIHUACAN
TOLTEC
AZTEC
TOTAL LESS (TEXCOCO BROWN AND TEXCOCO FABRIC-MARKED) 163 816 1203 1669 1090 1191 815 336 76 232 234 518 8368
TOTAL LESS TEXCOCO FABRIC-MARKED 360 3012 5668 5579 3071 3867 2307 809 199 465 699 750 32950
TOTAL SHERDS 360 3030 5720 5612 3208 5258 3957 3746 1311 2590 2012 2248 39002

FIG. 6. El Risco types. The grand total for all sherds less Texcoco Fabric-marked shown in the figure as 32,950 is in error. It should be listed as 24,582.

mens, systematically from bottom to top, in order to discern form qualities, especially rim shapes. While doing this form sorting, I was able to re-examine a few questionable categories and re-align some of the unallocated decorated sherds, either to types named by Tolstoy or an El Risco type. In aligning four types (San Martin Stamped, San Miguel Stick-Trailed, San Martin Corrugated, San Martin Punctated) I was guided by the type definitions instead of Tolstoy's allocation of these as Classic rather than Post-Classic types.

The final types derived by our study are listed graphically in figure 6 with three kinds of percentages illustrated. Brainerd (1951) has suggested eliminating the numbers of an extremely popular type when calculating such trends. Texcoco Fabric-marked is an obvious major type in the upper half of the deposit, attaining a maximum frequency of 85 per cent of all sherds in level 4. Figure 6 shows this type as a percentage of all sherds. Texcoco Brown is also shown as a percentage curve on this basis and very neatly demonstrates the complementary nature of the two curves.

We have thus dropped Texcoco Fabric-marked frequencies from the totals used in calculating the second curve shown for Texcoco Brown. This second curve shows a more gradual, decreasing trend toward the top levels. Because Texcoco Brown is also a very popular type and attains a maximum frequency of 73 per cent of all sherds in level XI, we have dropped *both* Texcoco Fabric-marked and Texcoco Brown from the totals in computing percentages of all other types.

In practice we drew up a chart showing all types but Texcoco Brown and Fabric-marked plotted in two ways: (1) using total of all sherds but Texcoco Fabric-marked as total sample; (2) using total of all sherds but Texcoco Fabric-marked and Texcoco Brown as total sample.

It is this latter percentage (2) which has been plotted on figure 6. While a complementarity was demonstrable for the curves of the two most frequent types, it was not seen in the two possible curves for the other types. We chose to show the second graph because it made trends more visible without altering range or modes.

POTTERY TYPES

With the exception of unnamed classes or "El Risco" types, our types conform to the descriptions set forth by Tolstoy (1958). Three main chronologically significant series are represented at El Risco.

Texcoco (Aztec), Tula (Toltec) and San Martin (Teotihuacan). Below (tables 1, 2) are listed the various types derived by Tolstoy for "Classic" and "Post-Classic" times.

Primary criteria for defining these types are surface finish, slip color, and decoration. Paste characteristics are essentially constant with the exception of the paste of "Thin Orange" and "Texcoco Fabric-marked." The paste charac-

TABLE 1

TOLSTOY'S POST-CLASSIC TYPES

A. Wares

I. Burnished

a. Monochrome Types

1) Texcoco Brown
2) T. Dark Brown
3) T. Black
} 16,835

4) T. White (39)
5) T. Incised (158)

b. Painted Types

1) Tula Red/Buff (840)
2) Mazapan Red/Buff (108)
3) Tula Brown/Brown (27)

II. Red

a. Texcoco Red (1025)
b. T. Black/Red (749)
c. T. Black & White/Red (96)
d. T. White/Red (21)
e. T. Yellow/Red (6)

III. Orange

a. Texcoco Orange (6,070)
b. Tenochitlan Black/Orange (663)
c. Tenayuca Black/Orange (214)
d. Tlatelolco Black/Orange (13)
e. Culhuacan Black/Orange (5)

IV. Plain

a. Undecorated (3,791)
b. Texcoco Moulded (23)
c. Texcoco Filleted (5)
d. Tula Red/Paste (10)
e. Heavy Incensarios (Tula) (12)

V. Brush-slipped

a. Tula Orange (306)
b. Tula White (115)
c. Tula Red-Orange (60)

VI. Fabric-marked

a. Texcoco Fabric-Marked (5,048)

VII. Tula Watercolored (18)

VIII. Plumbate and Metallic (8) (3)

TABLE 2

TOLSTOY'S CLASSIC TYPES

A. Wares

I. Thin Orange (854)

II. Burnished

II. Burnished

a. Monochrome Types (Undecorated)

1) San Martin Yellow-Brown
2) San Martin Brown
3) San Martin Dark Brown
4) San Martin Red-Brown
5) San Martin Black
6) San Miguel Orange

} 7,512

b. Monochrome Types (Decorated)

1) San Martin Incised (206)
2) San Martin Corrugated (23)
3) San Martin Stamped (18)
4) San Miguel Stick-Trailed (11)
5) San Miguel Zoned (6)

c. Red/Buff

1) San Martin Red/Buff (248)
2) San Martin Red/Buff Incised (53)

d. Polychrome and White/Red

1) San Martin Polychrome (18)
2) San Martin White/Red (33)

III. Red

1) San Martin Polished Red (974)

IV. Coarse Brown Ware

1) San Martin Coarse Brown (436)
2) San Martin Punctated (12)

V. Plain

1) San Martin Orange Plain (1,816)
2) San Martin Grey Plain (1,123)
3) San Miguel Plain (154) (Tentative Type)
4) Candeleros

teristics of the latter are a distinctive part of what Tolstoy calls the "Tepeyac" complex of pottery attributes. Since El Risco is in the center of the Tepeyac area we do indeed find considerable evidence for these distinctive pottery characteristics. They are most prominent in the type Texcoco Fabric-Marked but also appear in the paste of many of the Tula types. As a matter of fact, I have the distinct impression that the shape, basal break, rim form, and paste features of the Tepeyac complex first appear at El Risco in the Tula types and then continue in the Fabric-marked type.

I, too, have been impressed by the paste similarities among all of the types of both the Classic and Post-Classic periods. I have come to think of a "standard" paste with one other major group being the "Tepeyac" paste. These two are quite distinct and both exist in large numbers of sherds. The "Thin Orange" paste is quite distinctively different but appears in infinitesimal amounts at El Risco. Plainware paste is significantly more coarse than "standard"; Texcoco moulded paste is also coarser than "standard."

With these few qualifications in mind, it is possible to describe the types found at El Risco in the same terms as those used by Tolstoy. Figure 6 lists all the named types found at El Risco plus "El Risco Grey," "El Risco Engraved" and "El Risco Red/White," all of which are tentative types found at El Risco but not distinguished in the northern valley by Tolstoy. Seventeen other classes were defined in the course of our study but they were present in minute quantities and in no significant vertical distribution. They are: Black/Grey (I); White/Brown (II, XI, XII); Yellow/Brown (II); White/Buff/Red (V); Black/Brown/Red (VII); Black/Brown (VI); White and Black/Brown (VII); Black/White (VII); Brown/Grey (X); Polychrome (XI); Negative (X); Cuneiform (IV, VI, VII); Applique (V, VI, XI); Excised (VII, XI); Modeled (VII, XI, X); Aztec Glazed (II, IX); Punctate (VIII). Most were limited to one level (levels indicated in parentheses) and all were present in frequencies of one or two sherds. Modeled and Negative were found in frequencies of five each in level X.

Most, if not all, of these sherds may be considered as "trade" items from some source outside the immediate El Risco zone, or possibly color variants of the named types. The Polychrome, Negative, Excised, and Modeled are probably local types more characteristic of the ceremonial centers of the time and not usually found associated with rural household refuse.

Interpretation of Type Trends

Inspection of the curves shown in figure 6, and the table of percentage on which the curves are based (table 3), points up a number of significant trends. I have grouped my observations into three categories: (1) very good trend—this class has a well-defined continuous range, a well-defined single mode and is represented by a reasonably adequate sample; (2) good trend—this class has a well-defined and usually continuous range, a definite modality either singular or multiple and is represented by a reasonably adequate sample; (3) questionable trend—this class

has a questionable range, mode or sample, but some distinctive factor that suggests its possible value as a significant type. (Table 3 lists three kinds of information for each type per level: top number is actual numbers of sherds; middle number is percentage of total less Texcoco Fabric-Marked; bottom number is percentage of total less both Texcoco Fabric-Marked and Texcoco Brown. The type Texcoco Fabric-Marked is listed in terms of quantity and also percentage of total sherds. Texcoco Brown is listed in numbers, percentage of total less Texcoco Fabric-Marked and also as percentage of total of all sherds.)

1. *Very Good Trend*

Texcoco Moulded has a gap in its range at level IV but otherwise seems to show a very good trend with popularity peak in level II. This distinctive type apparently characterizes part of the ceremonial equipment of late Aztec times. Most of the sherds come from incensario bowls, perhaps similar in shape to the earlier Tula Red/Buff hollow handle ladle bowls. The excised sherds from level XII appear to be from a similar kind of bowl.

Tula Orange, Tula Red/Paste, and Mazapan Red/Buff all qualify with very good trends and also are associated vertically in the same series of levels (VI–X). Tula White and Tula Watercolored should perhaps be fourth and fifth members of this group even though ranges are not identical. All of these types are quite distinctive; Tula Orange and Mazapan Red/Buff are represented by the largest samples while Tula Watercolored and Tula Red/Paste are most distinctive. As an associated group these distinctive types apparently characterize the Toltec or Tula period occupation at El Risco.

San Martin Stamped is the only distinctive Classic type that has a very good trend at El Risco. Present in every level, except IV, it has an obvious peak of popularity in the bottom level. Because of this extensive vertical range it would seem wise to look for significant internal variations. The kind of stamping (style, size, position on the vessel) has been investigated in our sample and will be discussed below. This type evidently is the hallmark of the Classic period occupation at El Risco.

Texcoco Incised qualifies as having a very good trend but distinctions within the type must be clarified before suggesting its significance. From level IX upwards the majority of incised sherds are molcajete bases whereas below IX they are small bowl fragments with an incised exterior band or design and are quite similar to the vessel form and overall decorating mode of the type San Martin Stamped. While Texcoco Incised appears in every level but IV, its peak is in level XII and with the exception of a slight peak in level I appears to have a very good trend. Because there is a strong possibility that we have two distinct classes here under one type name, it is likely that the four lowest levels should more properly be called San Martin Incised (the Classic type). The incising styles here certainly conform to the rather loose definition set forth by Tolstoy. The trends in levels XII–VIII are parallel for both San Martin Stamped and Texcoco Incised. On this basis we may suggest that the association of the two types strengthens our evidence for a Classic period occupation at El Risco.

Tenayuca Black/Orange, the old Aztec II type, is present in very limited quantities at El Risco, but has a very good trend and is a distinctive type. It presumably characterizes the more elaborate or ceremonial expression of the earliest Aztec occupation at El Risco.

Texcoco Fabric-Marked certainly qualifies for this trend class in every way: it is very distinctive in both paste and surface finish; it has a well-defined range with well-defined, single mode and a most adequate size of sample. The type is apparently the hallmark of the Aztec period occupation at El Risco, but it comes into the picture along with Tula types so perhaps is also characteristic of the Tula occupation here. It should be said that the very few sherds of Texcoco Fabric-Marked found in levels IX, X, and XI were identified primarily on the basis of paste characteristics rather than surface finish. They were quite small in size and much eroded so that surface finish was not discernible in many cases.

2. *Good Trend*

Texcoco Black, Texcoco Red, and Tula Red/Buff all have good samples; range from top to bottom in the cut and are all bi-modal. Texcoco Black, in the lower levels, may be the Classic type San Martin Black and Tula Red/Buff may be San Martin Red/Buff. Texcoco Red, however, has both its peaks within the Aztec levels.

Texcoco Black/Red is not found in levels III

TABLE 3

EL RISCO SHERD COUNTS AND PERCENTAGES

	XII	XI	X	IX	VIII	VII	VI	V	IV	III	II	I	Levels
16214	*197* 54.7 54.7	*2196* 73. 72.4	*2465* 67.2 66.3	*3911* 70.1 69.7	*1981* 62. 64.5	*2676* 69.2 50.9	*1492* 64.7 25.	*473* 58.5 12.7	*119* 61. 4.7	*227* 49. 8.9	*245* 49. 12.2	*232* 31. 10.3	Texcoco Brown
1												*1* .1 .2	Texcoco Yellow/Red
39							*1* .04 .12	*2* .2 .6		*8* 1.7 3.4	*22* 4.4 8.7	*6* .8 1.2	Texcoco Moulded
15							*9* .4 1.1	*1* .1 .3		*2* .4 .8	*2* .4 .8	*1* .1 .2	Texcoco Black & White/Red
33					*3* .1 .3	*2* .05 .18	*16* .7 2.	*9* 1.1 2.7			*2* .4 .8	*1* .1 .2	Texcoco Black/Red
57						*7* .2 .59	*24* 1. 2.9	*9* 1.1 2.7	*3* 1.5 3.9	*1* .2 .4	*1* .2 .4	*12* 2. 2.3	Tenochtitlan Black/Orange
5		*1* .03 .12				*2* .05 .18	*1* .04 .12				*1* .2 .4		Texcoco White/Red
20				*2* .04 .12		*5* .13 .42	*12* .5 1.5	*1* .1 .3					Tenayuca Black/Orange
504			*11* .3 .9	*51* .9 3.06	*17* .5 1.5	*85* 2.2 7.1	*156* 6.8 19.1	*39* 4.8 11.6	*13* 6.6 17.1	*26* 5.6 10.9	*22* 4.4 8.7	*84* 11. 16.2	Texcoco Orange
155	*1* .28 .6	*11* .37 1.3	*14* .39 1.2	*27* .5 1.6	*5* .2 .5	*30* .8 2.5	*23* 1. 2.8	*17* 2. 5.		*5* 1. 2.1	*12* 2.4 4.7	*10* 1. 1.9	Texcoco Red
903	*13* 3.6 8.	*67* 2.2 8.2	*112* 3. 9.3	*232* 4.2 13.9	*192* 6.3 17.6	*91* 2.3 7.6	*57* 2.5 7.	*40* 5. 12.	*13* 6.6 17.1	*40* 8.6 16.8	*18* 4. 7.1	*28* 4. 5.4	Texcoco Black
1411	*63* 17.5 38.7	*125* 4.1 15.3	*161* 4.4 13.4	*199* 3.6 11.9	*219* 7.1 20.	*123* 3.2 10.3	*107* 4.6 13.1	*76* 9.4 22.6	*16* 8.2 21.	*54* 11.6 22.7	*58* 11.6 22.8	*210* 28. 40.5	Plain
835	*5* 1.4 3.	*83* 2.8 10.2	*164* 4.5 13.6	*198* 3.5 11.9	*63* 2. 5.8	*131* 3.4 11.	*38* 1.6 4.7	*40* 5. 12.	*16* 8.2 21.	*34* 7. 14.3	*28* 5.6 11.	*35* 5. 6.8	Tula Red/Buff
29		*9* .3 1.1	*5* .1 .4	*1* .02 .06	*2* .07 .2	*6* .15 .5	*2* .08 .24		*2* 1. 2.6	*1* .2 .4	*1* .2 .4		Tula Brown/Brown
473			*18* .5 1.5	*140* 2.5 8.4	*143* 4.7 13.1	*121* 3.1 10.2	*51* 2.2 6.3						Tula Orange
28			*5* .1 .4	*7* .12 .4	*14* .5 1.3	*1* .03 .08	*1* .04 .12						Tula Red/Paste
95			*8* .2 .7	*35* .6 2.1	*37* 1.2 3.4	*7* .2 .59	*8* .3 .98						Mazapan Red/Buff

	XII	XI	X	IX	VIII	VII	VI	V	IV	III	II	I	Levels
46			*30* .8 2.5	*3* .05 .18	*9* .3 .8	*4* .1 .34							Tula White
26		*5* .17 .6	*5* .1 .4	*3* .05 .18	*9* .3 .8	*1* .03 .08	*3* .1 .37						El Risco Engraved
39		*1* .03 .12	*19* .5 1.6	*5* .09 .3	*3* .1 .3	*3* .08 .25	*7* .3 .86					*1* .1 .2	Texcoco White
55		*3* .1 .37	*24* .65 2.	*22* .4 1.3	*4* .1 .4	*2* .05 .18							Tula Watercolored
14		*2* .07 .25	*2* .05 .17	*1* .02 .06	*9* .3 .8								El Risco Metallic
2			*1* .03 .08		*1* .03 .09								Culhuacan Black/Orange
3026	*42* 11.7 25.8	*328* 10.9 40.2	*489* 13.3 40.6	*670* 12. 40.2	*337* 11. 31.	*532* 13.8 44.7	*272* 11.8 33.4	*89* 11. 26.5	*11* 5.6 14.5	*62* 13. 26.	*72* 14.4 28.3	*122* 16. 23.5	El Risco Grey
26	*2* .5 1.2	*7* .23 .86	*7* .19 .58	*1* .02 .06	*1* .03 .09	*1* .03 .08	*1* .04 .12		*1* .5 1.3	*2* .4 .8	*1* .2 .4	*2* .3 .4	El Risco Red/White
236	*8* 2.2 5.	*61* 2. 7.5	*51* 1.4 4.2	*49* .9 2.9	*16* .5 1.5	*23* .6 1.9	*10* .4 1.2	*6* .7 1.8		*2* .4 .8	*8* 1.6 3.1	*2* .3 .4	Texcoco Incised
228	*25* 6.9 15.3	*103* 3.4 12.6	*61* 1.7 5.	*13* .23 .78	*2* .07 .2	*5* .13 .42	*8* .3 .98	*5* .6 1.5		*1* .2 .4	*3* .6 1.2	*2* .3 .4	San Martin Stamped
18			*4* .1 .33	*7* .12 .4	*4* .1 .4		*3* .1 .37						San Miguel Stick-Trailed
2		*1* .03 .12					*1* .04 .12						San Martin Corrugated
1				*1* .02 .06									Thin Orange
3	*1* .28 .6		*2* .05 .17										San Martin Punctated
14420		*18* .6	*52* 1.4	*33* .6	*137* 4.3	*1391* 26.5	*3650* 61.	*2937* 78.	*1116* 85.	*2075* 82.	*1513* 75.	*1498* 67.	Texcoco Fabric-Marked
41	*3*	*9*	*10*	*1*	*1*	*9*	*4*	*2*	*1*		*1*		Not Listed
	XII	XI	X	IX	VIII	VII	VI	V	IV	III	II	I	
24582	*360*	*3012*	*3668*	*5579*	*3071*	*3867*	*2307*	*809*	*195*	*465*	*499*	*750*	Total Less Texcoco Fabric-Marked
8368	*163*	*816*	*1203*	*1668*	*1090*	*1191*	*815*	*336*	*76*	*238*	*254*	*518*	Total Less Texcoco Fabric-Marked and Texcoco Brown
39002	*360*	*3030*	*3720*	*5612*	*3208*	*5258*	*5957*	*3746*	*1311*	*2540*	*2012*	*2248*	Total Sherds

and IV, but otherwise has a good trend with a peak popularity early in the Aztec period.

Texcoco Orange has a range starting in the Tula levels and three modes in the Aztec levels. Interestingly enough the earliest mode corresponds to the level of the mode for Tenayuca Black/Orange. The second mode occurs in the same level as the mode for the Tenochtitlan Black/Orange. Tenochtitlan Black/Orange (Aztec III) is confined to the Aztec levels but has a suggestion of bi-modality. Main peak is in level IV, but level I frequency suggests a second mode. El Risco Grey is present from top to bottom of the cut with a broad peak from level VIII to XI. There is a suggestion of bi-modality but this type evidently is a basic type for the entire time of Cut A.

El Risco Metallic, El Risco Engraved, and San Miguel Stick-Trailed, are all listed in this category since they are very distinctive. The Metallic type develops an increasing trend from level XI to level VIII, where it is apparently chopped off at its peak. This type is evidently a local imitation of the widespread Meso-American Plumbate ware. El Risco engraved has a double mode and generally shows an increasing frequency towards the bottom of the deposit. San Miguel Stick-Trailed is the only definite Classic period type to be found in this category. It shows a fairly consistent frequency from level VI to X. All these types, however, are found in very small absolute quantities.

TABLE 4

EVALUATED EL RISCO TYPES

Distinctive Types

Aztec

** Texcoco Moulded
** Tenayuca Black/Orange
** Texcoco Fabric-Marked
* Texcoco Red
* Texcoco Orange
* Texcoco Black/Red
* Tenochtitlan Black/Orange
Texcoco Black and White/Red

Tula

** Tula Orange
** Tula Red/Paste
** Tula White
** Tula Watercolored
** Mazapan Red/Buff
* El Risco Metallic
* El Risco Engraved
Culhuacan Black/Orange
Texcoco White

Teotihuacan

** San Martin Stamped
* San Miguel Stick-Trailed
* (San Martin Incised)
(San Martin Red/Buff)

Continuous Types

** Texcoco Incised (San Martin Incised in lower levels)
* Texcoco Black
* Texcoco Red
* Tula Red/Buff (San Martin Red/Buff in lower levels)
* El Risco Grey
El Risco Plain
Texcoco Brown

* Good trend established at Cut A, El Risco II.
** Very good trend established at Cut A, El Risco II.

3. *Questionable Trend*

El Risco Plain is present in all levels but shows a peak frequency at both I and XII. Texcoco Brown is one of the two main types at El Risco. Present in large numbers in every level it has a generally decreasing trend towards the top levels, a broad mode from level VIII to XI. Texcoco Black and White/Red is present in small numbers from level VI to I with a vaguely decreasing trend from a peak at VI. It apparently is restricted to and thus distinctive of the Aztec period. Culhuacan Black/Orange is known from our cut at El Risco by only two sherds, one in level VIII and one in level X. As a representative of the former Aztec I type we find no trend but a presence restricted to the time of Tula types.

Texcoco White has a fair trend from level XI to VI with a peak at X. This distribution suggests that the type is not an Aztec type, but is rather most characteristic of the Teotihuacan Period. Here we differ with Tolstoy's findings.

4. *Other Types*

A number of the other types present at El Risco have interesting distributions in the light of the above information. One Thin Orange sherd occurred in level IX where it can be considered either Toltec or Teotihuacan in period. The Modeled, polychrome, Brown/Grey, White/Brown, San Martin Punctated and San Martin Corrugated all appeared in the Teotihuacan period. The latter two, at least, tend to support our suggestion of a Teotihuacan period.

In table 4 we have listed a summary evaluation of the types found in our cut at El Risco. The "distinctive" types are those which in combination represent the "diagnostics" for the period.

The "continuous" types are those which appear as background for all three periods and must be considered in specific combinations with "distinctive" types for periodization to be complete.

FORM ANALYSIS

Because sample sizes are generally rather low, I have not grouped the various trends of change in form that were analyzed. Table 5, however, listing the form class frequencies gives a good idea of what those changes are.

Interpretation of Form Trends

Because the various attributes of form here considered are not equally comparable we assume that the levels have standard, comparable samples and thus depend upon percentages based on vertical rather than horizontal totals. These percentages have not been listed in table 5 which simply presents frequencies.

1. *Comales* (sample is 598). Basal fragments are consistently about three times as numerous as rims. The general frequency trend is good, ranging through all levels with peak frequency at VI. Below level VII the bases are characterized by a coarse-grained, "crackled" texture. The fine grained "sandy" rough bases start in level VIII and characterize the bases from level VII up. This change in basal surface texture is apparently correlated with a change in rim form frequency (see table 6), as A_5, A_6, A_7, A_8, comal rims are associated with coarse-grained bases. The overall trend appears to be one of minor frequency in the lower five levels (but peak in XI), a sudden increase in frequency to a major peak in level 6 and a more gradual diminution to level I.

2. *Handles* (sample is 145). All samples here are inadequate. There is a fair suggestion of a trend in the categories loop, split loop, and strap. All forms are most common in levels VI–XI; loops show the best trend with a peak at X. Double and split loops show a (specious?) disconformity, but the split loops definitely precede double loops. Lugs have inconclusive spread but cluster around level VI; stroop (a combination of strap and loop forms) seems to be a fair intermediate category in terms of its distribution.

3. *Supports* (sample is 218). Ringstands, hollow cylindrical and solid nubbin feet have the best samples and distribution. Ring stands are characterized by a definite but discontinuous trend with a strong peak building toward the

TABLE 5—EL RISCO FORM CHANGES

Levels	Bases	Comales: Rims	Comales: Total	Handles: Loop	Handles: Lug	Handles: Dbl. loop	Handles: Lug	Handles: Split loop	Handles: Stroop	Handles: Strap	Handles: Total	Feet: Hollow: Ring stand	Feet: Hollow: Slab	Feet: Hollow: Spider	Feet: Hollow: Cylinder	Feet: Hollow: Irregular	Feet: Hollow: Bulbous	Feet: Hollow: Mammiform	Feet: Solid: Mammiform	Feet: Solid: Perf. hemis.	Feet: Solid: Slab	Feet: Solid: Cylinder	Feet: Solid: Bulbous	Feet: Solid: Conical	Feet: Solid: Nubbin	Feet: Solid: Total	Ladle handles: Hollow cylinder	Ladle handles: Trough	Ladle handles: Solid conical	Ladle handles: Flattened conical	Bases: Right angle	Bases: Obtuse angle	Bases: Tepeyac	Bases: Molcahete	Vessels: Jar	Vessels: Carinated bowl	Vessels: Olla	Vessels: Total
I	15	7	22	2		2				1	5	3			1										4		5			1	5			1			12
II	38	15	53	2							2	1													1		3							1		7
III	32	5	37	2		1	1			1	4				1										3	1	2									4
IV	17	3	20	1							2	2													2		2									2
V	64	15	79			1	1		7	2	3	2		1	3		4				1	1	2	6	23	1	2			1	1					8
VI	146	60	206	3		2			6	9	20	5		1	15	1	2					1		11	37	5	3			12	4	21	1	1			72
VII	81	37	118	2						6	9	4	1		11	1	6		1		1	1	2	17	28	9	4		2	4	11	41	4	2	1		43
VIII	3	7	10	1						15	34	11		2	11		1	2		1	2		3	2	51	7	23	16	4	7	13	21	18	15	9	6	74
IX	9	1	10	4		10		7	5	1	27	18	1	2				1		2			2	1	31	4	15	1	4	13	6	9	6	2	1	1	50
X	6	7	13	4				6	1		18	25	1		1		1								34	1	1									41
XI	11	11	22	12							2	1					1								2		1							2		6
XII	6	2	8					2			2	1																								1
Total	428	170	598	54	5	16	2	13	15	40	145	71	2	4	44	2	15	2	1	1	5	5	1	12	218	34	77	13	2	8	54	33	20	17	1	382
	Bases	Comales	Comales	Handles	Handles	Handles	Handles	Handles	Handles	Handles	Handles	Feet	Feet	Feet	Feet	Feet	Feet	Feet	Feet	Feet	Feet	Feet	Feet	Feet	Feet	Vessels	Ladle handles	Ladle handles	Ladle handles	Ladle handles	Bases	Bases	Bases	Bases	Vessels	Vessels	Vessels	Vessels
				Loop	Lug	Dbl. loop	Lug	Split loop	Stroop	Strap	Total	Ring stand	Slab	Spider	Cylinder	Irregular	Bulbous	Mammiform	Mammiform	Perf. hemis.	Slab	Cylinder	Bulbous	Conical	Nubbin	Total	Hollow cylinder	Trough	Solid conical	Flattened conical	Right angle	Obtuse angle	Tepeyac	Molcahete	Jar	Carinated bowl	Olla	Total

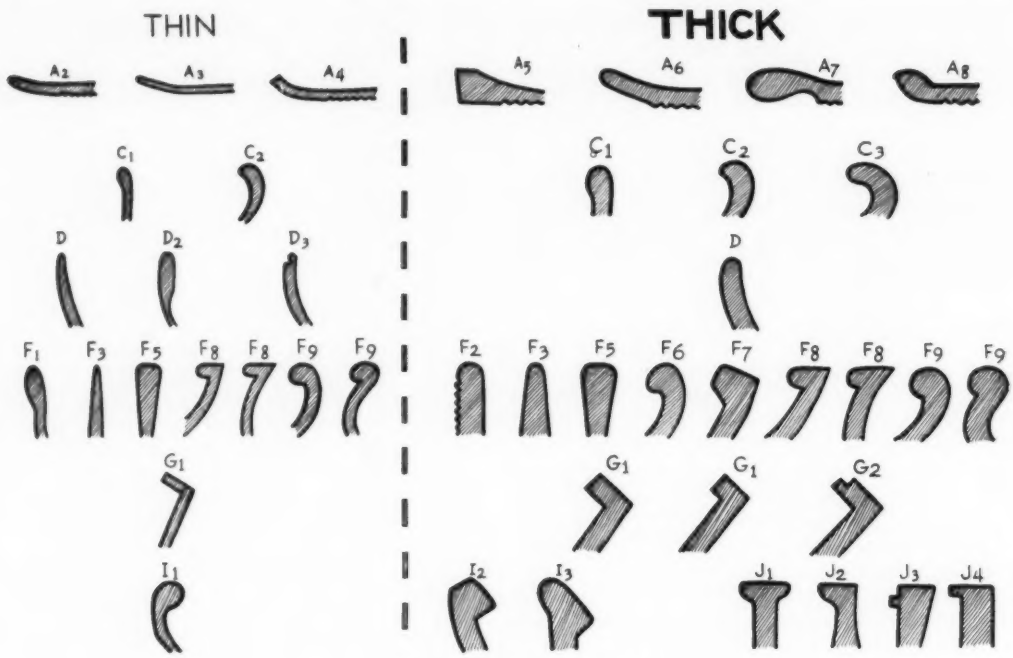


FIG. 7. El Risco rim modes.

bottom levels. Hollow cylinder legs also have a spotty distribution, but seem to have a fair trend with a strong peak in level VII. Solid nubbin feet have a continuous range from VI to XI, are characterized by a good trend with a peak at IX. Slab legs occur only below level V; mammiform only in X.

4. *Vessel Form* (sample is 364). Ladles and bases appear to give the most significant trends. The hollow cylinder ladle handles show a good trend from II to XII with a peak at level VIII. Shoehorn or "trough" handle ladles are about twice as frequent overall as the cylinder handle form. The trough ladles are absent only in level IV, have a trend skewed to an increase in the lower levels with a peak frequency in level X. The "Tepeyac" style rounded base ranges mainly from V–XI with a good trend and a peak at VII. The obtuse angle base is distributed essentially the same way but with a trimodal curve (VI, IX, XI). Right angled bases increase towards the bottom. Molcajetes in level VI and above are Aztec in style, those below are Toltec. Storage jars have an inconclusive trend, but tend to cluster in the richest levels at VI. Carinated bowls are found in level VI and below and increase toward the bottom.

RIM ANALYSIS

Examination and study of all rim sherds led to the development of a number of modes or form classes, which were derived independently of Tolstoy's similar modes. El Risco modes were then compared with the total range of form Tolstoy considered. He did not distinguish as many varieties in Post-Classic as in Classic materials. We have taken over his designations where applicable but do not completely duplicate his modes. In addition we considered the factor of thickness and developed separate modes for each of the two categories "thick" and "thin." Sample size suggested that we abandon this, however, for final interpretation. We have thus combined thick and thin rims into one chart (figure 7) to illustrate the rim modes.

Interpretation of Rim Mode Changes

Five modes account for 80 per cent of all rim forms. In decreasing order the most frequent

modes are as follows:

D, C_1, F_3, A_2, C_2, G_1, F_2, F_8, F_1, and C_3.

Six modes have what we may call a "good trend" on the basis of range, frequency and peak. These are A_2, A_3, A_4, C_1, F_2, and F_3. In addition five modes have a "fair trend." They are C_2, and D, F_1, F_8, and G_1. Comparison with Valley of Mexico collections at the American Museum of Natural History and the University of California (Los Angeles), as well as large rim sherds from El Risco, suggests the following vessel form correlates for several of the rim modes.

comales—A_2, A_4–A_8
saucers—A_3
bowls—D, thin C, thin F_1, F_3
jars—thick C, F_4–F_7, F_9, G, J
apazte—F_2, F_3

Inspection of the frequency distribution (percentage of all rims per level indicated for each rim mode underneath actual frequency) of all El Risco rims shown in table 6 suggests that certain modes are temporally distinctive and others are not. Distinctive forms are:

A_2—Tula comal
A_3—Teotihuacan saucer
A_4—Aztec comal
A_5–A_8—Teotihuacan comal
F_1—Tula bowl (especially Tula Orange)
F_2—Tula apazte (Tepeyac complex)
G_1—Tula jar (especially Tula Red/Paste)

Rim modes with continuous distribution are:

C_1—Tula peak?
C_2—Tula peak?
C_3—Teotihuacan peak?
D—Tula peak
F_3—Tula peak (Tepeyac apazte)

Interpretation of Texcoco Brown Rims

Because the type Texcoco Brown was the majority type in all but Aztec levels, the rims of this type were analyzed separately, in order to point up the role of this basic utilitarian type. Table 7 lists the frequency distribution (percentage of all rims per level indicated for each rim mode underneath actual frequency) of all Texcoco Brown rim modes. Here again a few extremely popular modes account for most of the rims. The most popular are D, C_1, F_3, A_2, C_2, F_2, G_1, and C_3. The "good trend" category is represented by A_2, A_3, A_4, C_1, D, F_2, and G_1. A "fair trend" is shown by A_6, A_8, F_1, F_3, F_9, and I_2.

The same vessel forms are represented by this sample, but the distinctive and continuous modes differ somewhat from the larger sample. Continuous modes are the same with less certain suggestions for peak frequency. The distinctive modes repeat except for F_1 which is essentially restricted to Tula Orange and Texcoco Fabric-Marked types.

In the analysis of Texcoco Brown rims the thin-thick distinction was made systematically in order to test impressions that thickness increased toward the base of the cut. Analysis did not bear this out completely. Modes listed under "thin" were A_2, A_3, A_4, C_1, C_2, D, F_1, F_3, F_5, F_8, F_9, G_1, and I_1. Modes listed under "thick" were A_5, A_6, A_7, A_8, C_1, C_2, C_3, D, F_1, F_2, F_3, F_5, F_6, F_7, F_8, F_9, G_1, G_2, I_2, I_3, J_1, J_2, J_3. Plotting of Texcoco Brown frequency distribution in this way demonstrated that comal forms A_2 and A_4 were always thin while comal forms A_5–A_8 were always thick; saucer form A_3 is thin; there is a strong tendency for C_1, C_2 and especially D forms to be thin; there is a strong tendency for F_2, F_5, F_6, F_7, F_8, F_9, G, I_2, I_3 and all J modes to be thick; F_3 is predominantly thin but becomes approximately 50 per cent thick in the bottom two levels.

No rigid criteria of measurement were adopted for making this thin-thick distinction but a dividing line of about $\frac{3}{16}$ inch would generally result in a similar division. Throughout the cut a ratio of about three thin for every thick sherd would probably hold with a slight tendency for the thicker sherds to increase in frequency in the lower levels.

SPECIAL DATA

The stamped (Tolstoy, 1958: fig. 8, *v* and *w*), incised (Tolstoy, 1958: fig. 8, *a–l*), and engraved sherds form a group of materials susceptible of detailed stylistic analysis, but our sample is inadequate for significant correlations here. Our quantified observations of this material do suggest some trends, however.

Three varieties of stamping were distinguished on the basis of primary design motif: checker, rectilinear, and curvilinear. In the four lowest levels curvilinear patterns are most popular, ranging from 46 to 63 per cent; rectilinear pat-

TABLE 6

EL RISCO RIM CHANGES

Rims	A_2	A_3	A_4	A_5	A_6	A_7	A_8	C_1	C_2	C_3	D	D_2	D_3	F_1	F_2	F_3	F_5	F_6	F_7	F_8	F_9	G_1	G_2	I_1	I_2	I_3	J_1	J_2	J_3	J_4	Total
Levels																															
I	1 2		6 12					3 6	5 10		17 35			1 2		9 18					7 14										49
II			15 30					7 14	2 4		17 34	1 2				2 4				3 6	2 4		1 2							50
III			5 12.5					6 15		4 10	18 45					3 7.5	2 5	1 2.5		1 2.5										40
IV	3 11.5							5 19			15 58					2 8						1 4								26
V	6 6		6 6		3 3			12 12	4 4		41 40	4 4				12 12	1 1	5 5	2 2	2 2	2 2				1 1	2 2					103
VI	56 17				2 .6		2 .6	36 11	17 5	3 1	124 38			12 4	8 2	41 13	1 .3		2 .6	5 1.5	3 1	7 2		1 .3	2 .6	1 .3					323
VII	34 7						3 .6	79 16	13 2.6	5 1	214 43	2 .4	2 .4	22 4	24 5	69 14	1 .2			10 2		14 3				4 .8		1 .2	1 .2		498
VIII	7 2.5							59 21	15 5.4	1 .4	128 46			13 5	5 2	22 8		5 2		11 4		13 5					1 .4				280
IX		1 .2		1 .2				130 23	5 .9	11 2	282 50	2 .4		7 1.2	14 2.5	54 9.5	18 3		1 .2	7 1.2	4 .7	24 4	1 .2	1 .2	2 .4	2 .4					567
X		2 .5			7 2		2 .5	81 19	29 7	6 1.4	178 42				13 3	45 11	3 .7			16 4	3 .7	20 5	1 .2		9 2	4 1			2 .5		421
XI		4 1		1 .3	6 1.5	.3 .8	1 .3	74 19	14 4	12 3	194 50	1 .3				43 11	1 .3	5 1	11 3	4 1	1 .3	2 .5			8 2		2 .5				387
XII		2 2				1 2	1 2	11 23	2 4		29 60				1 2	2 4						1 2									50
Total	107	9	32	2	18	4	9	503	106	42	1,257	10	2	55	65	304	27	16	16	59	22	82	2	3	22	13	3	1	1	2	2,794

TABLE 7

TEXCOCO BROWN, ALL RIMS

Rims	A_2	A_3	A_4	A_5	A_6	A_7	A_8	C_1	C_2	C_4	D	F_1	F_2	F_3	F_4	F_6	F_7	F_8	F_9	G_1	G_2	I_1	I_2	I_4	J_1	J_2	J_3	Total
Levels																												
I			4											4					4									12
			33											33					33									
II			14						2		4							2	1									23
			61						9		17							9	4									
III			1					2		4	8			1	2													18
			6					11		22	44			6	11													
IV	1							2			5			1						1								10
	10							20			50			10						10								
V	5		6					1	3		16			2		4	1		2				1					41
	12		15					2	7		39			5		10	2		5				2					
VI	42				2			19	11	3	64		8	15			1		9			1	2					177
	24				1			11	6	2	36		5	8			.5		5			.5	1					
VII	26						3	34	13	5	109		24	46				6		11				3		1	1	282
	9						1	12	5	2	39		9	16				2		4				1		.3	.3	
VIII	7							19	13	1	68	2	3	8		4		9		7								141
	5							13.5	9	.7	48	1	2	6		3		6		5								
IX				1				71	3	11	167		11	29	17			4	3	20		1		1				339
				.3				21	1	3	49		3	9	5			1	1	6		.3		.3				
X		2			7		2	35	21	6	81		13	23				9	3	10	1		9	3				225
		1			3		1	16	9	3	36		6	10				4	1	4	.4		4	1				
XI		1		1	5	3	1	46	11	12	109			33	1	5	11			1			8		1			249
		.4		.4	2	1	.4	18	4	5	44			13	.4	2	4			.4			3		.4			
XII		2				1	1	5	1		12		1	1						1								25
		8				4	4	20	4		48		4	4						4								
Total	81	5	25	2	14	4	7	234	78	42	643	2	60	173	20	13	13	30	22	51	1	2	20	7	1	1	1	1,542

terns range from 34 to 46 per cent; checker patterns range from 3 to 6 per cent. Checker and curvilinear suggest early popularity peaks and rectilinear a later peak. Most stamped vessels are small bowls of the D rim mode, some are C_1 mode and a very few are F_3 mode. The stamped design usually appears as a narrow band encircling the bowl near the base. There is some suggestion of larger stamps in materials from the upper levels.

The incised category comprises two main classes of sherds. In the upper levels (above IX) most of the sherds are molcajete bases with the incising as an interior design or scoring. In the four lowest levels incised sherds increase greatly in frequency and, except for the Toltec molcajete bases in level IX, are fragments of small bowls with D or less frequently C_1 rims. The incised lines are narrow in about 90 per cent of the cases, broad or trailed in 10 per cent. Design is most frequently a simple meander enclosed by parallel straight border lines, the whole appearance being as a band encircling the bowl on the outer wall, very much like the band of stamping on San Martin Stamped.

We have distinguished engraved sherds from incised on the basis of time of execution. Engraving may be considered as "post-firing incising." Distributed only from level VI–XI, this technique is limited to Tula and Teotihuacan types and increases in frequency toward the lowest levels. The engraved designs are most similar to the incised outer bands, but also occur as borders for a red rim band on Tula Red/Buff sherds. One example in IX is a Toltec molcajete base.

Texcoco Fabric-Marked is a most interesting and unusual type that will be considered in detail in a separate article. Tolstoy has discussed it at some length, concluding tentatively that it is of special use in connection with salt production. We will add our comments to this in the interpretive section.

Our attempt carefully to control the movement of every sherd from cut to type file was seriously thwarted by one drastic loss. The upper five levels were excavated in two units (north, south) but for level IV only the north unit arrived in Pittsburgh. This was the only known loss of specimens. Some of the minor distribution gaps and the small sample size in level IV can thus be attributed to the missing one-half of the cut at this level.

FIGURINES

In figure 8 are shown the recognizable fragments ($\frac{2}{3}$ natural size) of figurines, "adornos," and other small modeled ceramic pieces from the excavation.

Levels X, XI, and XII each contained a number of adorno fragments, presumably many of them from Teotihuacan period incensarios like those illustrated by Tozzer (1921: pl. 14) and Linné (1934: figs. 149, 150). Because of the general lack of controlled excavation information on Classic and Post-Classic figurines we here merely present and briefly describe a selection of our finds. The entire complex from our El Risco cut will be examined in detail at a later date when more adequate comparative data are available.

The items in figure 8 represent five main kinds of items: (1) human representations (face or full figure); (2) animal heads; (3) decorated pot fragments; (4) stamp fragment; (5) grotesque faces.

In figure 8, A is a very stylized small animal head from level I. B is a crudely molded, flat-backed (pregnant female?) human torso with a suggestion of ear plugs, formal hair treatment, and possible jutting headpiece (level V). C is the head and shoulders region of a flat-backed mold-made human figure (level VII). D is a stamp-like (with protruding handle on back) fragment of mold-made human figure (level VII). E is the rim fragment of a small pot with applique nose and "coffee bean" eyes. Very much like the decorated pot found with burial 1 (level VII). F is a crocodile-like animal snout made of the typical Tepeyac paste (level VII). G is the corner of a rectangular stamp (level VII). H is the only human face from level VIII; stylistically it is very similar to I. Traces of red paint remain. I is a stylized human face with rectilinear headgear covering the forehead as in H (level IX). J is a different human face with similarities to the standard Teotihuacan III face as illustrated in Vaillant, 1944: pl. 21, bottom, center and Linné, 1934: fig. 178 (level IX). K, L are both small facial fragments. K is apparently similar to H and I, L is like C (level IX). M, N, O, and P are all animal heads from level IX. M is a long stylized snout; N is evidently a serpent with extended tongue; O is a deer and P an ocelot. Q is a complete human figure (except for the feet), flat-backed and mold-made. The face is like the standard Teotihuacan III form, but the scale-like torso covering treatment

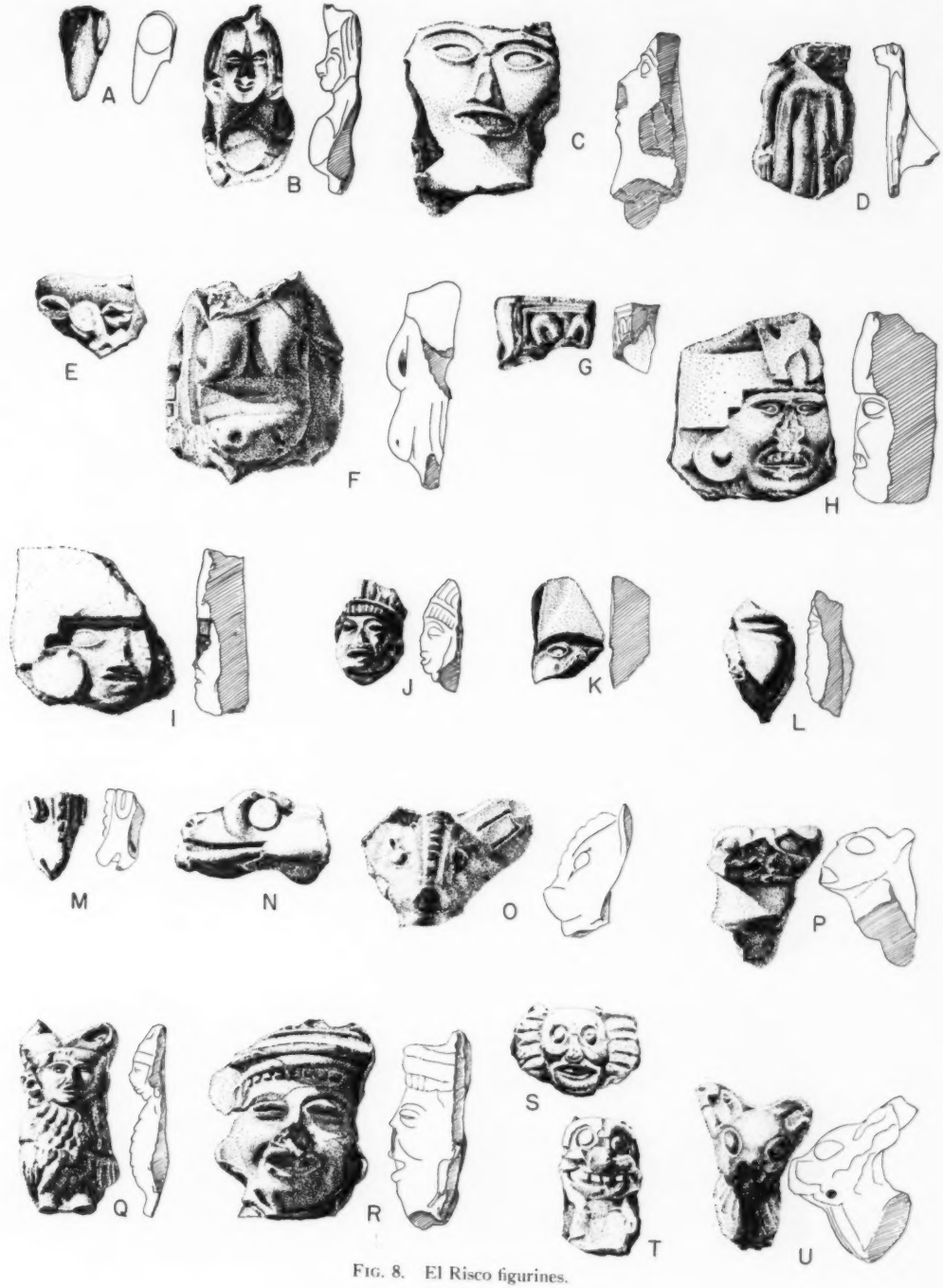


FIG. 8. El Risco figurines.

TABLE 8

Level	Items	Total
I	2 proj. points; 3 cuchillos; 5 flakes; 1 chert flake	11
II	2 cuchillos; 2 worked flakes; 6 flakes	10
III	3 cuchillos; 2 worked flakes; 8 flakes	13
IV	None	
V	3 cuchillos; 4 worked flakes; 5 flakes	12
VI	7 cuchillos; 1 scraper; 1 flake; 1 chert flake; 3 basalt flakes; 1 basalt chopper	14
VII	9 cuchillos; 7 flakes; 1 large hafted scraper	17
VIII	6 cuchillos; 11 flakes; 1 small stemmed point-reamer	18
IX	6 cuchillos; 1 ripple-flaked blade; 1 worked flake; 16 flakes; 1 mano (tufa)	25
X	2 side-notched points; 1 corner-notched point; 1 drill base; 2 reamers (on blades); 28 cuchillos; 4 worked flakes; 32 flakes; 1 chert flake	71
XI	11 cuchillos; 3 worked flakes; 24 flakes; 7 basalt flakes; 2 chert flakes	47
XII	5 flakes	5
		237

is reminiscent of the Xipe Totec figures as illustrated by Linné, 1934: fig. 113 (level X). R is a flat-backed, but semi-round face with special head-dress broken away. It is heavily coated with red paint (level X). S and T are grotesque faces, S evidently being a skull with winged extensions and T a wide-eyed, bared-tooth form. Both are covered with red paint (level X). U is a black slipped bowl rim with exterior protruding stylized deer head from level X.

We have had very little success in tracking down specific parallels for most of these forms. An attempt will be made to do this more systematically in a future paper. Tozzer (1921) has illustrated a number of similar forms in his plate 13. Here, his forms B and C compare with our O and U. Our T is similar in style to his A.

NON-CERAMIC DATA

Table 8 lists the various stone items found in our excavations at El Risco II. Unless otherwise indicated, the stone is obsidian and it is worked by chipping.

It is interesting to note that overall frequencies are quite low, with a definite maximum in the Teotihuacan period. The major artifact type is the *cuchillo* or flake knife; sometimes worked along one or two edges or an end, and always showing signs of some use. The most distinctive projectile points came from level X; they are illustrated (natural size) in figure 9.

PERIODIZATION

From the distribution of specific pottery types through the deposit there are definite indications of three major clusters or time periods. The distribution of various elements of pottery form also suggest this three-period alignment. In the chart which follows, the details of these two kinds of distribution are integrated into period characteristics, presenting both quantitative and qualitative evidence for the existence of these three major time periods at El Risco II.

At the same time that we emphasize the cultural elements distinctive of the three periods it should be remembered that there are at least two kinds of evidence which suggest continuity from top to bottom of the deposit: the continuously present though quantitatively variable elements; the degree of cultural overlapping between contiguous periods of time.

Our original period terms were Texcoco, Tula, and San Martin to correlate with the major pottery series names. Generalization to the

FIG. 9. El Risco projectile points.

terms Aztec, Toltec, and Teotihuacan formalizes our assumptions about broader cultural or civilizational relationship involved. While we recognize the inherent dangers in this "ethnic" label approach we here make explicit our assumptions without trying to document and support them in specific detail. Such a task is outside the scope of the present paper.

Periodization Elements

AZTEC PERIOD (levels I–VII)
(I) Natural Elements: location in upper (slanted) strata
(II) Cultural Elements:

Pottery Types

Texcoco Moulded—incensarios are distinctive, late peak
Texcoco Incised—molcajetes are distinctive
Tenayuca Black/Orange—distinctive of early Aztec period
Texcoco Fabric-Marked—very distinctive even though overlapping with Tula period
Texcoco Red
Texcoco Black/Red—distinctive, early peak
Texcoco Orange—major characteristic, Tula overlap
Tenochtitlan Black/Orange—two peaks
Texcoco Black and White/Red—limited to this period

Pottery Forms

Vessels:

Comales of Texcoco Brown or Texcoco Orange with A_4 (see fig. 7 for drawings of all rim modes) rim and fine-grained, "sandy" roughened base characterize the period, obtaining peak popularity late in time.

Molcajetes of Texcoco Orange or one of the Black/Orange types are distinctive of the early part of the period. The incised "grater bowl" pattern is different from that of the Tula molcajetes. Slab legs are characteristic.

Apaztes of Texcoco Fabric-Marked with F_1 and F_2 rims are very distinctive attaining a peak frequency in the middle part of the period.

Incensarios of Texcoco Moulded are very distinctive for the period, attaining a peak frequency late in time.

Ladles are present in small frequencies; the trough handle type in Texcoco Brown is more common than the hollow cylinder handle type in Tula Red/Buff.

Handles:

Some loop and some strap handles from this period, but very low in frequency. The lug or unperforated handle made from a double loop appears only in this period while the simple loop-lug overlaps with Tula. All are most frequent in the early part of the period.

Feet:

The nearly complete absence of feet during this period is the most significant fact of our El Risco distribution.

Rims:

A_4 rims show the most significant distribution; they are limited to the Aztec period with a peak frequency in level II.

TOLTEC PERIOD (levels VI–X)
(I) Natural Elements: located between the horizontal strata and stone concentration around burial
(II) Cultural Elements:

Pottery Types

Tula Orange—especially bowls with F_1 rims
Tula Red/Paste—especially jar with G_1 rim
Mazapan Red/Buff
} very distinctive; same distribution; peak in VIII

Tula White—early peak
Tula Watercolored—early peak
El Risco Metallic—peak at VIII
El Risco Engraved—declines from peak in Teotihuacan (?)
Culhuacan Black/Orange—limited to Toltec period

Pottery Forms

Vessels:

Comales of Texcoco Brown, El Risco Grey, Texcoco Orange, or Texcoco Black with A_2 rim and "sandy" roughened base is characteristic of late Toltec times, but overlaps into early Aztec period.

Bowls of Tula Orange with F_1 rim, flat bottoms and solid nubbin feet are the most distinctive markers for this period. Maximum absolute frequency of small Texcoco Brown bowls with D rim.

Apazte form in Texcoco Brown with F_2 rim attains peak popularity.

Jars with G_1 rim of Tula Red/Paste type (usually very small vessel) reaches peak relative and absolute frequency in this period.

Molcajetes of Texcoco Brown or Tula Red/Buff with distinctive incised patterns and "spider" leg characterize the period and attain early peak.

Ladles—Hollow, cylindrical handle type (often of Tula Red/Buff) attains peak frequency in this period. Texcoco Brown trough handle type more abundant but has declined from San Martin peak frequency. "Solid" and "flattened" conical ladle handles are incensario lid fragments as illustrated by more complete Portesuelo specimens seen in Los Angeles.

Handles:

Double loop and strap handles of Texcoco Brown type each have peak in Toltec period.

Feet:

Hollow cylindrical legs characterize the period as do solid nubbin feet. The nubbin feet have an early peak while the hollow legs have a late peak.

Bases:

The Tepeyac base characterizes the period and is found most frequently represented on Texcoco Brown and Tula Orange types. Peak frequency occurs in level VII.

Rims:

F_1 and F_2 modes are most distinctive of the period with J_2, J_3 and possibly I_3, as associated distinctive types. The A_2 comal rim is most frequent and distinctive comal style, but A_6 and A_8 are present.

TEOTIHUACAN PERIOD (levels IX–XII)

(I) Natural Elements: located beneath the level of stone concentration

(II) Cultural Elements:

Pottery Types

San Martin Stamped—in form of small bowl with D or C_1 rims in ratio of 3:1. Peak in level XII.

Texcoco Incised—in levels IX–XII in form of small bowls with D or C_1 rims. Conforms to San Martin Incised type.

San Miguel Stick-Trailed—distinctive type present in small amounts.

El Risco Engraved—increases in frequency toward bottom level.

El Risco Plain—marked increase in bottom two levels.

Texcoco White—distributed through both Teotihuacan and Toltec periods with peak in XI.

San Martin Punctated—distinctive, very minor element.

San Martin Corrugated—distinctive, very minor element.

Thin Orange—distinctive, very minor element.

El Risco Modeled, Negative and Polychrome types—rare, but apparently characteristic of the Teotihuacan period.

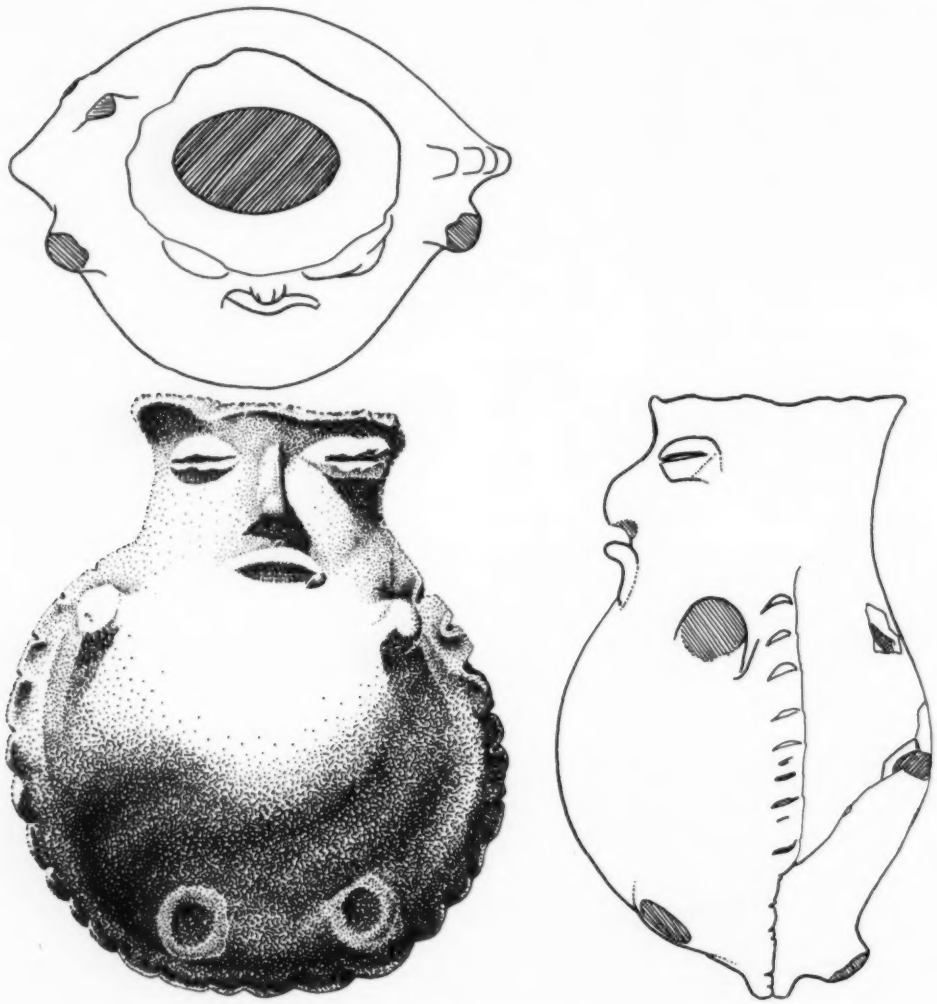


FIG. 10. El Risco burial pot.

Pottery Forms

Vessels:

Comales of Texcoco Brown with A_5–A_8 rims and, rarely, loop handles on lip. Small numbers, but maximum frequency in XI.

Saucers of Texcoco Brown with A_3 rim.

Bowls—small, "grape-fruit half" of Texcoco Brown with stamped, incised or engraved and rarely red/buff decoration; larger, wide-mouth Texcoco Brown with ring stand base, rarely red/buff; carinated with marked shoulder.

Ladles of Texcoco Brown in trough or "shoehorn" handle form reach peak frequency. Hollow cylindrical handle ladles present.

Handles:

Loop handles have peak in San Martin period, both single and split (pseudo double).

Feet:

The only mammiform feet occur in the Teotihuacan period. Slab feet are at maximum frequency.

Bases:

Ring stand bases occur from top to bottom of the deposit but achieve marked frequency peak in level XI. Right angle bases also are at maximum in XI, but other forms are still common.

Rims:

Maximum frequencies for C_1, C_2, F_5–F_8, I_2, J_1 and J_4 modes.

BURIAL DESCRIPTION

As luck would have it, our stratacut was practically centered on a burial found at the 10-foot level. During excavation of the 9 to 10-foot level we encountered a number of large rocks especially in the north one-half of the cut along the east wall. First indication of the burial was this concentration of stones. Second indication was a sinking hollow sensation as a pick cut into a small clay bowl. Trowel work uncovered the freshly broken fragments of the bowl and a complete small effigy-face jar (shown natural size in fig. 10). Both of these were placed near the skull of a tightly flexed skeleton which partly protruded into the 9–10-foot level but was essentially contained in the 10–11-foot level. Cleaning of the 10-foot floor exposed the outlines of the burial. The next level was taken down except for a block containing the skeleton which was then cleaned and removed separately (figs. 11 and 12). The skeleton was that of an adult, tightly flexed, head to the south, facing east. The two pots shown in figure 11 were found approximately nine inches southeast of the skull. They were removed along with the soil and rocks which covered the burial.

FIG. 11. Burial No. 1 at El Risco, discovery.

FIG. 12. Burial No. 1 at El Risco, partially cleared.

No evidence for a pit was observed; the scattered covering of stone may be fortuitous but it was a feature concentrated above and around the burial and not present elsewhere in the cut. Our working method in the level of discovery was with a small hand pick. Our practice of chopping a cut of about 10 inches and then skin-

ning off the final 2 inches is responsible for disturbing the only definite association of artifacts (the two pots); but is also responsible for preserving the skeleton intact.

Study of the skeleton has not yet been accomplished. The effigy jar is very much like one found by Linné at Teotihuacan, presumably in a Mazapan grave (Linné, 1934: figs. 109, 110).

COMPARISONS

Tolstoy's dissertation is the most complete, up-to-date consideration of ceramic problems in the Valley of Mexico. Because of this fact as well as our joint field work and subsequent discussions and correspondence, it will be most useful to concentrate on his data and conclusions for comparisons.

CERRO PORTESUELO

One other major unit of information, however, is the two-season excavation project carried out at Cerro Portesuelo (in the southeastern part of the Valley of Mexico) by the late George W. Brainerd. Five main sources of information about the Portesuelo project are now available to me:

Sources of Information

1. Brainerd published a brief statement (in "Notes and News," *American Antiquity* **31** (4): 441) regarding his 1955 work at two specific locations—several levels of domestic architecture with Toltec period remains, a second excavation in a sequence of building substructures pierced by intrusive burials. He described the occupation as beginning in Teotihuacan times.

2. A conference on the Valley of Mexico was called by Brainerd and held at the Harvard Peabody Museum on November 16, 1955. Here Brainerd stated that his work had been excavations in a habitation area of "Coyotlatelco—Toltec—Aztec 3" and in a separate ceremonial area. He described the domestic area as being composed of Toltec, Aztec 2 and 3 with Red on Buff pottery (either San Martin or Tula) most characteristic of the lower levels. Specific finds of what he called Coyotlatelco complete vessels were found in graves with classic Teotihuacan types of pots. These were described and illustrated with color slides. At this same conference, Pedro Armillas suggested the usefulness of the concept of a folk-urban continuum for interpretation of the Valley; he suggested that at least three kinds of units could be distinguished: folk—peasant—urban. Important factors here to be considered are the degree of population concentration and the actual growth of the population. Problems discussed at this time included: consideration of Coyotlatelco or red on buff pottery generally as a peasant pattern; the question of functional differentiation of sites and material lots; the question of evidences for irrigation and chinampas.

3. Frederic Hicks (MS), a student of Brainerd, prepared a paper entitled "The Chronological Position and Cultural Significance of Coyotlatelco Pottery of the Valley of Mexico." This is based largely on the first (1954) season at Portesuelo. This study suggested that the popularity maximum and final stages of this style were fairly well defined but the initial appearance was still a question. The association of a ring-based bowl form in this style with the earliest Portesuelo contexts suggested a Classic period time level. Several problems were recognized, especially the one of distinguishing the Coyotlatelco style from Tolstoy's San Martin Red on Buff and the problem of the Portesuelo graves with Teotihuacan II vessels associated with red on buff vessels identified as Coyotlatelco. Hicks sums up as follows (p. 10):

> The available data, then, suggest that Coyotlatelco is essentially a Toltec pottery type, which was probably introduced at about the time of the destruction of the Teotihuacan ceremonial center (possibly earlier), gradually replacing Classic pottery types in the Valley of Mexico, and reaching a peak in popularity during the period of Tula's ascendancy in the Valley. At some time, probably about the time of the fall of Tula or shortly thereafter, it began a decline in popularity, finally disappearing altogether in early or middle Aztec times, having been replaced by Aztec pottery types.

4. In the last week of December, 1956, and the first week of January, 1957, I traveled to Los Angeles on an American Philosophical Society grant and spent seven days examining the Portesuelo materials at the University of California. While there I completed five main kinds of activity:

A. I observed the synoptic Portesuelo exhibits Brainerd had arranged:

i. Early Classic was shown by Teotihuacan II and III pots and figurines.
ii. Late Classic was shown by Teotihuacan IV pots, figurines, and incensarios plus the

first Coyotlatelco round bowls, ring-base bowls, and hollow cylindrical handle ladles.

iii. Early Post-Classic was shown by pots of later Coyotlatelco, Mazapan, plumbate and "gingerbread" figurines.

iv. Late Post-Classic was shown by pots of Aztec III, IV styles plus mold-made figurines, spindle whorls, and pottery stamps.

B. I talked with two of the students who had helped with the laboratory work on the 1954 materials: J. Handler had sorted the types of sherds on the basis of Brainerd's original categories; P. Kunkel who had done the original 1954 calculations.

C. I examined a small collection (less than 100) of whole pots or large fragments which were on open storage in the archaeological lab workroom.

D. I studied in detail the classified trays of sherds stored in room 309. This was all 1954 data, and only a small fragment of the total, but was assumed to be representative of the entire lot.

E. No field notes were available. There was a blueprint map of the 1954 digging locations, a series of laboratory notes on sorting classes and a preliminary tabulation of the 1954 sorting. From discussions with Kunkel I learned that the excavation units were 5 meter squares dug in 30 cm. levels. Brainerd had sorted away all "formless" and undecorated sherds in the field. These items left in Mexico probably account for from 60 to 90 per cent of the total excavated. Thus 10 to 40 per cent per level returned to the University of California at Los Angeles.

5. I have a statement from Brainerd in a letter to Tolstoy (a carbon had been sent to me). He had completed the second season of work at Portesuelo and suggested the Boston conference, having just received a copy of Tolstoy's manuscript. He arranged for all three of us to meet after the main conference to discuss details of our respective projects. Brainerd had selected the Portesuelo site at Tolstoy's suggestion as a starting place to investigate the Teotihuacan-Tula transition period. On November 1, 1955, he wrote:

> . . . Cerro Portesuelo has proved a good choice. We worked there two months this summer with 40 men and 3 graduate students. . . . A year ago we got about 150,000 sherds and a few whole pieces. They have all been typed and tabulated, but few summaries and illustrations have been made. The tabulation is in over 100 categories, with major emphasis on vessel form and decoration. This year our sherd count has not yet been made, but must range several times that of last year; we brought home nearly 3 tons of material after discarding perhaps 6–8 tons of sherds. We also have about 100 whole or easily restorable pots, plus quite a welter of other artifact material, architecture, etc. As you may guess we are pretty well swamped. Time range of the occupation is Teotihuacan II–Aztec III with much Teotihuacan IV–V and Toltec material.

In preparing for the conference at Harvard, Brainerd formalized his chronological ideas in an outline of periods as follows:

I. Classic Teotihuacan (Teoti. II, III, Miccaotli, Xolalpan)
II. Late Classic (Coyotlatelco, Ahuitzotla-Amantla, Teoti. IV–V)
III. Tula-Toltec (Mazapan-Aztec I, Plumbate horizon, Culhuacan B/O)
IV. Aztec (Aztec II–IV, Tenayuca, Tenochtitlan and Tlatelolco B/O)

These categories agree very well with the headings used in the synoptic exhibits but the placing of Coyotlatelco only in Late Classic and including Aztec II in the final category perhaps indicates Brainerd's feeling that both Late Classic and Tula-Toltec could be seen as rather discrete entities, clearly separable from either Classic Teotihuacan at early levels or Aztec in later levels. In the light of Tolstoy's analysis, the big question is about the nature of the relationships between Late Classic and Tula periods.

Our Study

Most of my time with the Portesuelo collection was spent in going over the type file in room 309, Haines Hall, checking these classes against the more completely integrated types I had already derived at El Risco and adjusted to Tolstoy's types. I found that very few of the sherds were unfamiliar and the proportions present were most like those present in the lowest levels at El Risco. Brainerd's emphasis on form classes was different from our own and it is here that most discrepancies arose. From the 109 sorting classes he used for the 1954 material I have selected 43 that can be compared directly with our El Risco data. I abstracted the distribution information about these 43 classes from Kunkel's final tabulation of the 1954 data and in table 9 present the figures for seven of the trenches. These seven were selected as having the most adequate

TABLE 9

PORTESUELO SORTING CLASSES

FORMLESS (discarded)	62 A	62 B	62 C	62 D	62 E	62 F	62 G	62 H	62 I	62 J	62 K	60 A	60 B	60 C	60 D	60 E
	31	57	68	44	65			48	46	35	15	350	302	212	134	53
1. Subrim flange plates			1													
2. W/ upcurve lip (A_2, A_4)	1												2			1
3. Incurved lip, parallel sided flange (A_7)	1														2	1
4. Upcurved rims (A_6)													2		1	
5. Upturned rims (A_8)	1															
6. Triangular short flange (A_5)					1		3		1		1	1	2	2	2	1
7. Nearly imperceptible flange (I_2)					3				2			1	1		2	2
8. R/O bowl rims red rim banded													1			
9. R/O bowl body des int. bottom												1		1		
10. R/O bowl rim hem. C.I.D.			1													
11. R/O bowl body hem. C.I.D.													1			
12. R/O bowl rim hem. dull color matte C.I.D.			1									3	3	1		
13. R/W or. var. hem. bowl rim	1											1	2	1		
14. R/W body & legs			1	1							1	1				
15. R/W bowl rims pl. & stripe rim hem																
16. R/W bowl rims pl. & red rim stripe & flare												3				1
17. Small hollow dome etc. legs																
18. Hollow slab legs	1											3	11	1	2	2
19. Dome legs, small dark bowls												1	1			
20. Hollow cyl. bowl legs												9	3	1		
21. Thin dark flare rims, beakers incised		1	1													1
22. Trough handle dipper			2	1								2	1	1	1	
23. B/R rims & body														3	1	1
24. Cylinder dipper handles					1							1		1		2
25. Incised and grater bottoms							1					1		1		
26. Flare rim platters	1								1		1	2		1		
27. Platters w/ outbent or outcurve lip									1							
28. Platters rims dir. lip int. slip only																
29. Platters rims dir. lip slip on both sides																
30. B/O bowl rims fine line & dot																
31. B/O bowl rims heavy line zacate bord.											1					
32. B/O beaker heavy line																
33. B/O beaker rims heavy line diag. hatch bord.																
34. B/O bowl body fine line & dot																
35. B/O slab leg																
36. B/O beaker fine line & dot																
37. Mold pressed design flare rim vessel															1	1
38. Incised dichrome ext. design flare rim bowls													1	1		
39. Mold pressed bands on flare rim bowls													1			
40. Mold pressed bands on rounded bowls																
41. Dichrome w/ mold pressed design												1	1	1	1	
42. Incised dichrome on rounded bowls												1	2		1	
43. Ring bases to hem. bowls		1		1					1							
TOTAL	6	2	7	3	5	—	4	1	5	—	4	31	35	16	14	12

TABLE 9—(*Continued*)

PORTESUELO SORTING CLASSES—(*Continued*)

	38 A	38 B	38 C	38 D	38 E	38 F	37 A	37 B	37 C	37 D	37 E	37 F	37 G	37 H	37 I	37 J	35 A	35 B	35 C	35 D
FORMLESS (discarded)		58				175	119									14	3030	2396	135	213
1. Subrim flange plates																	26	13		
2. W/ upcurve lip (A_2, A_4)																	9	15		
3. Incurved lip, parallel sided flange (A_7)																	6	9		
4. Upcurved rims (A_6)																	3	4		
5. Upturned rims (A_5)						1											6	1		
6. Triangular short flange (A_8)							1										3	4		
7. Nearly imperceptible flange (I_2)																	15	6		
8. R/O bowl rims red rim banded	1																16	6		
9. R/O bowl body des int. bottom							1										11	4		
10. R/O bowl rim hem. C.I.D.							1										11	5		
11. R/O bowl body hem. C.I.D.																	9	3		
12. R/O bowl rim hem. dull color matte C.I.D.							1	1									5	3		
13. R/W or. var. hem. bowl rim																	10	5		
14. R/W body & legs							1										36	40		
15. R/W bowl rims pl. & stripe rim hem.																	20	18		
16. R/W bowl rims pl. & red rim stripe & flare								1									1			
17. Small hollow dome etc. legs																				
18. Hollow slab legs																	2	1		
19. Dome legs, small dark bowls																	1		1	
20. Hollow cyl. bowl legs										1							1	1		
21. Thin dark flare rims, beakers incised																	9	11		
22. Trough handle dipper							2										29	13		
23. B/R rims & body			1														4	11		
24. Cylinder dipper handles								1									17	11		
25. Incised and grater bottoms							1										19	15		
26. Flare rim platters							2										14	5		
27. Platters w/ outbent or outcurve lip							1										7	3		
28. Platters rims dir. lip int. slip only							1	1									22	8	1	
29. Platters rims dir. lip slip on both sides																	24	11		
30. B/O bowl rims fine line & dot																	18	11		
31. B/O bowl rims heavy line zacate bord.								1									14	17	1	
32. B/O beaker heavy line																	9	12		
33. B/O beaker rims heavy line diag. hatch bord.																	19			
34. B/O bowl body fine line & dot																	6	1		
35. B/O slab leg																	24	13		
36. B/O beaker fine line & dot																				
37. Mold pressed design flare rim vessel																	1	1		
38. Incised dichrome ext. design flare rim bowls																				
39. Mold pressed bands on flare rim bowls																				
40. Mold pressed bands on rounded bowls																				
41. Dichrome w/ mold pressed design							1										3	1		
42. Incised dichrome on rounded bowls																		3		
43. Ring bases to hem. bowls																				
TOTAL	1	—	1	—	—	1	13	5	—	1	—	—	—	—	—	—	430	285	3	—

TABLE 9—(*Continued*)

PORTESUELO SORTING CLASSES—(*Continued*)

FORMLESS (discarded)	29 A	29 B	29 C	29 D	29 E	29 F	29 G	25 A	25 B	25 C	25 D	25 E	25 F	25 G	24 A	24 B	24 C	24 D	24 E	24 F	24 G
	787	410	310	190	73	42			170	698	236		97	61	124	169	289	123		234	15
1. Subrim flange plates																					
2. W/ upcurve lip (A_2, A_4)										2						2					
3. Incurved lip, parallel sided flange (A_7)	4	1	1	2	2															1	
[4. Upcurved rims (A_6)	2	2																			
5. Upturned rims (A_8)		1									1										
6. Triangular short flange (A_5)	1	1						1									1	1	1		
[7. Nearly imperceptible flange (I_2)	1						1														
8. R/O bowl rims red rim banded	5			1	2		2				1			1			3	1		1	
9. R/O bowl body des int. bottom	4	3	1	2	1		2	2		4											
10. R/O bowl rim hem. C.I.D.		2	1	4			1														
11. R/O bowl body hem. C.I.D.	1	1	1	1	1								1								
12. R/O bowl rim hem. dull color matte C.I.D.		1			1		1	1													
13. R/W or. var. hem. bowl rim	1																				
14. R/W body & legs	3	1	2	2						2	1				1		1				
15. R/W bowl rims pl. & stripe rim hem.	7	9	3	4	1		4		1							1					
16. R/W bowl rims pl. & red rim stripe & flare	6	1	2	1					1											1	
17. Small hollow dome etc. legs	1	1																			
18. Hollow slab legs																					
19. Dome legs, small dark bowls	7	1	1										1								
20. Hollow cyl. bowl legs		1																1	3		
21. Thin dark flare rims, beakers incised	1	1								1	1					1	6			1	
22. Trough handle dipper	1	2		4	1			2		6	4									1	
23. B/R rims & body	1	1						1			1						1			1	
24. Cylinder dipper handles	5	1																		1	
25. Incised and grater bottoms	2	2	1	2			1			5			3				1	1		1	
26. Flare rim platters	10	1	1						1	3	4		3			1		2		1	
27. Platters w/ outbent or outcurve lip				3				1	2				1								
28. Platters rims dir. lip int. slip only				1	1						1						1				
29. Platters rims dir. lip slip on both sides	4			2	1		3										1				
30. B/O bowl rims fine line & dot	1																				
31. B/O bowl rims heavy line zacate bord.	2	3																			
32. B/O beaker heavy line	2	1	1			1															
33. B/O beaker rims heavy line diag. hatch bord.	1																				
34. B/O bowl body fine line & dot	1						1														
35. B/O slab leg	1																				
36. B/O beaker fine line & dot	5	3	1							1								1		1	
37. Mold pressed design flare rim vessel																					
38. Incised dichrome ext. design flare rim bowls	2								1	1										2	
39. Mold pressed bands on flare rim bowls										1											
40. Mold pressed bands on rounded bowls																					
41. Dichrome w/ mold pressed design																					
42. Incised dichrome on rounded bowls	1												1								
43. Ring bases to hem. bowls	2																			1	1
TOTAL	85	41	16	29	11	1	16	8	6	26	14	—	10	1	1	5	15	7	4	14	1

TABLE 9—(*Continued*)

PORTESUELO SORTING CLASSES—(*Continued*)

	49 A	49 B	49 C	49 D	49 E	49 F	49 G	49 H	49 I	41 A	41 B	41 C	41 D	41 E	41 F	41 G	41 H	41 I
FORMLESS (discarded)	140	99	95	168	50	221	88	168	275	73	112	102	130	299	968	469	241	145
1. Subrim flange plates														1	1			
2. W/ upcurve lip (A_2, A_4)	2	1				1		1		1	2	2	2		4	1	1	
3. Incurved lip, parallel sided flange (A_7)				2	2	1			1							1	1	1
4. Upcurved rims (A_6)										1							1	
5. Upturned rims (A_8)	1	3		2	3	5			1		1		2	3	9	2	1	
6. Triangular short flange (A_5)															1			
7. Nearly imperceptible flange (I_2)	1	1		1	1	1		2				1		1	2		5	
8. R/O bowl rims red rim banded				1				1								2	2	2
9. R/O bowl body des int. bottom																		
10. R/O bowl rim hem. C.I.D.																		
11. R/O bowl body hem. C.I.D.																		
12. R/O bowl rim hem. dull color matte C.I.D.																		
13. R/W or. var. hem. bowl rim																		
14. R/W body & legs															1			
15. R/W bowl rims pl. & stripe rim hem.						1			2							1		
16. R/W bowl rims pl. & red rim stripe & flare	1		1												2	2	1	
17. Small hollow dome etc. legs						1										1	1	
18. Hollow slab legs								1		2					3	3		
19. Dome legs, small dark bowls														8	9	10	3	1
20. Hollow cyl. bowl legs						1	2		2		1	2	1	4	24	1		
21. Thin dark flare rims, beakers incised	2		1	5		4			4	1			3	4	24	2	6	
22. Trough handle dipper			1	1	1	2	2	1	4		1				2			
23. B/R rims & body	1	1												1	16	17	1	
24. Cylinder dipper handles									1	1		1		1	2	2	2	1
25. Incised and grater bottoms													2	1	1	1	5	
26. Flare rim platters	1	3				1									3		2	
27. Platters w/ outbent or outcurve lip								1	1				2			1		
28. Platters rims dir. lip int. slip only	2	1		1	1	1		2	1									
29. Platters rims dir. lip slip on both sides											1							
30. B/O bowl rims fine line & dot																		
31. B/O bowl rims heavy line zacate bord.												1						
32. B/O beaker heavy line																		
33. B/O beaker rims heavy line diag. hatch bord.																		
34. B/O bowl body fine line & dot																		
35. B/O slab leg																		
36. B/O beaker fine line & dot														1	1	1		
37. Mold pressed design flare rim vessel																	1	
38. Incised dichrome ext. design flare rim bowls	1				1			1							1			
39. Mold pressed bands on flare rim bowls	1														1			
40. Mold pressed bands on rounded bowls							1						1				1	
41. Dichrome w/ mold pressed design											1							
42. Incised dichrome on rounded bowls	1	2											1	2	9	3	2	
43. Ring bases to hem. bowls					1	1		1	2									
TOTAL	14	12	3	13	10	20	5	11	19	6	7	7	14	27	116	51	36	5

samples, but as can be seen the frequencies are quite low in absolute terms. (Trenches are numbered 24, 25, 29, 35, 41, 49, 60. Levels are listed alphabetically, "A" being the top level. Trench 35 was a major area of expansion and concentration in the 1955 season.)

Brainerd's sorting classes can be compared to those used at El Risco as follows:

1–7 are comal forms (Brainerd's class of "sub-rim flange plates").

- 1 is the general class.
- 2 is equivalent to our A_2 and A_4 rim modes.
- 3 is equivalent to our A_7 rim mode.
- 4 is equivalent to our A_6 rim mode.
- 5 is equivalent to our A_8 rim mode.
- 6 is equivalent to our A_5 rim mode.
- 7 is equivalent to our A_5 or I_2.

8, 9 are Red/Buff color types, either Tula or San Martin.

10–12 are classic or "wavy-line" Mazapan Red/Buff.

13–16 are Mazapan Red/Buff (orange/cream variant).

17–20 are comparable to our "foot" categories.

21 is comparable to our C_1 and C_2 rim modes but more restrictive since incised.

22 is same category in both schemes.

23 is Texcoco Black/Red.

24 is same category in both schemes.

25 is our class "molcajetes" plus?

26–29 are our F_5–F_9 rim modes.

30–36 are Aztec Black/Orange types.

- 30 is Tenochtitlan Black/Orange.
- 31 is Tenayuca Black/Orange.
- 32 is Culhuacan Black/Orange (or occasionally Tenayuca).
- 33 is Tenayuca Black/Orange.
- 34 is Tenochtitlan Black/Orange.
- 35 is ?.
- 36 is Tenochtitlan Black/Orange.

37, 39–41 are San Martin Stamped.

38, 42 are part of Texcoco Incised from El Risco lower levels (i.e., San Martin Incised).

(The 66 classes deleted from Brainerd's tabulation were essentially complete vessel or rim form categories of the color types listed in our table 9.)

Information from Kunkel and pencil notations on the field map indicate that trench 35 was the area of best stratigraphy. Expansion here in 1954 was made by trenches 29, 38, 39, plus 96 in 1955. Trench 62 was also considered a prospect for good stratigraphy. Units 24 and 25 are labeled as "good" on the map, perhaps as loci of Coyotlatelco pottery.

Brainerd developed whole vessel form classes which largely determined the sorting classes. They were:

- plates—our comales
- basins—our thick rims of F, G, I and J rim modes
- platters—deep, comal-like forms with F and G rim modes
- jars—thick D and G rim modes
- beakers—thin C, D and F rim modes
- bowls—thin B, C, D and F_1 or F_2 rim modes.

The bowls and beakers in whole vessel form are quite distinct. Flat-based, flare-rim dark beakers with right angle base are the basic "Classic-looking" ware; round-based Black/Orange shouldered beakers are the basic Aztec form. Bowl forms most distinctive are the hemispheric and ring-base styles.

Basically, our two sets of data are quite different. Samples are generally more adequate at El Risco whether discussing form or surface color or decorated types. El Risco provides better stratigraphic evidence also in terms of thickness of occupation. At Portesuelo, apparently, time ranges of from Teotihuacan II to Aztec III are found compressed into deposits only 5–8 feet deep. The 30 cm. levels would thus appear to be too thick to show detailed vertical changes. An additional factor of difference is that *all* sherds are included at El Risco whereas intensive selection for form and decoration has occurred at Portesuelo.

Specific comparisons with the Portesuelo chart are, however, useful and instructive in several categories.

Pottery Types

Red/Buff—Distribution with peak frequencies in the upper levels of trenches 35, 29 suggests predominance of Tula Red/Buff.

Mazapan Red/Buff—Distribution with peak frequencies in upper level of trench 35 suggests that excavation levels are too gross to detect difference between this type and Tula Red/Buff. Trench 29 distribution suggests that upper 5 levels are within Tula period.

San Martin Stamped—Distribution is based on very low frequencies, but still suggests range back into Classic, as in trench 60 where it predates dipper peaks and is associated with early comales, platters and Red/Buff forms. Similar distribution is found

in trench 41. (One specimen of San Martin Stamped with red-painted zones occurred at El Risco in level XII; also at Portesuelo in level G in trench 49 and levels D and H in trench 41.)

Pottery Forms

Comales—Trench 35 data tend to support our evidence for a late Tula and late Aztec peak for A_2 and A_4 modes, as well as an earlier peak for the A_5–A_8 modes. Trench 29 also supports the temporal spread of A_5–A_8 modes. Both trenches 49 and 41 support the general picture of A_5–A_8 distribution, especially the early peaks of A_7 and A_5. Trenches 29, 41, and 49 seem to be purely in pre-Tula times judged by comal distribution alone.

Trough Handle Dipper—Distribution in trench 29 with peak at level of earliest Mazapan parallels El Risco. In trench 49, a peak in lowest level associated with a peak of A_5–A_8 comales is parallel to El Risco. The trench 41 peak in middle levels associated with A_5 and A_7 peaks is parallel to El Risco and suggests trench 41 lowest levels are earliest at the site.

Cylinder Handle Dipper—Distribution in trench 60 suggests peak at top level, later than trough handle peak and thus parallel to El Risco. This is also true of the distribution in trench 29. Distribution in trench 41 suggests a peak one level lower than the trough handle peak. This is contrary to the trend at El Risco and may be related to functional factors (e.g., use as early grave goods).

Ring Bases—Only present in the second level of trench 35 and the lowest two levels of trench 24. Range in lower levels of trench 49 with peak at bottom. Best distribution in trench 41 with peak parallel to trough dipper peak but range going earlier. This similar to El Risco distribution.

On the basis of these specific comparisons with El Risco perhaps we can make some general guess as to the time span involved in the various Portesuelo cuts.

Trench 24 (7 levels) apparently Tula to late Classic, but thin.

Trench 25 (7 levels) apparently Tula to late Classic, but thin.

Trench 29 (7 levels) ranges from Tula at top to late Classic (?) at base. Seems to be heavily Mazapan even to lowest levels.

Trench 35 has Aztec, Tula and Teotihuacan periods all compressed into the upper two of its four levels. Has largest sample.

Trench 37 and 38 are too unproductive for usefulness.

Trench 41 (9 levels) Early Tula to Classic with fair sized sample.

Trench 49 (9 levels) Early Tula to Classic with good sized sample.

Trench 60 (5 levels) apparently starts in the Tula period and goes back into the Classic.

Trench 62 (11 levels) apparently goes back well into the Classic period but is too small a sample for usefulness.

Determinants for Classic in the above guesses have been the factors used to establish the Classic period at El Risco plus the presence of platters as found at Portesuelo. This seems to be a fair judgment since the majority of factors used by Tolstoy to define Classic cannot be determined from Brainerd's data. In selecting features from the 1954 Portesuelo tabulation I did not include Brainerd's categories for dark bowls (4 classes), thin dark beakers (3 classes)—I did include "thin dark flare rim beaker, incised"—basin rims (3 classes) and a number of other categories. These, especially, appear to be the Classic hallmarks and should have been included. In the study time available at Los Angeles, however, I concentrated on the factors of similarity to El Risco that could be most easily ascertained.

Interpreting Brainerd's ideas on the basis of my limited study of the Portesuelo collections, the general nature of Classic vessels from the general digging at Portesuelo is expressed in terms of dark color (grey and black) for the beakers with sharp basal ridge (right angle) and often nubbin feet. Vessel form tends to be cylindrical but flaring; thin to moderately thick rims are common. These are our C_1 and C_2 categories. Dark colored platters and basins with F_5–F_8, I and J rims are also apparently an emphasis in Classic. To these general characteristics can be added the emphasis on ring base bowls. Incidentally, a fair proportion of the ring bases from 1954 which I examined were Thin Orange ware, although not classified as such. This is another factor giving the lower levels a general Classic flavor.

In addition to these elements from the general digging are the burial caches with Teotihuacan II and III well-stylized forms associated with Coyotlatelco.

We have similar evidence for this growing "classicity" in the lower levels of El Risco in terms of the frequency patterns for C_1, C_3, F_5–F_8, I_2, J_1, and J_4; the frequency of El Risco Grey and Texcoco Black; the level XI peak for right angle and also obtuse angle bases.

From my study and handling of the sherds at UCLA I assured myself of the correlations listed above regarding Brainerd's sorting classes and our El Risco types. I also saw some of the rare

but distinctive San Martin and Tula types in a tray marked "non-rim rare wares." These items were examples of San Miguel Stick-Trailed, Tula Watercolored, and Tula Red/Paste. From trays marked "orange ware" I sorted out a number of Tula Orange specimens and also some Thin Orange. The majority here were Texcoco Orange. All of these types were set aside and labeled with my identification as to type. It was not possible to get provenience data for these sherds but they are indicative of both Tula and Classic contexts. The presence of Thin Orange is particularly significant because of its significance as a Classic marker with a peak frequency in Xolalpan (Teotihuacan III).

I also scanned but did not make notes on the few burial lots (?) available. My noted comment, "Various burial pots are Classic-like" now strikes me as singularly cryptic. At the time I was more concerned with sherd information and was unable to get many data on the burials. These grave associations are important, however, as Hicks has suggested. While it is possible that some of the vessels are heirlooms it seems more likely that we have distinctive grave styles combined with more common or utilitarian styles.

Interpretation of Site

It is quite difficult to evaluate and integrate the fragmentary information now available from Portesuelo. Both the nature of the site itself as a community or settlement pattern and the chronological value of the sherd material are still largely unknown. Our brief study and analysis has indicated some support for the fragmentary interpretive statements made by Brainerd.

Structurally, Portesuelo is seen as a large and complex habitation area characterized by both domestic and ceremonial architecture. The latter is apparently not of impressive size or nature but it has been distinguished from "habitation" areas. The sherd refuse from habitation areas can roughly be dated by comparison with Tolstoy's survey data and the finds at El Risco. A time span ranging from Early Classic to Aztec is indicated, but the major occupation is apparently in Late Classic and early Post-Classic times. As an architectural and settlement zone the site is apparently comparable to Azcapotzalco.

Brainerd was strongly influenced by Vaillant's interpretation of Azcapotzalco as "Late Classic" or Teotihuacan IV, V, and saw, I believe, a parallel between the two sites. I have no concrete evidence for this other than his periodization and his exhibits. The Teotihuacan II and III remains in burials are an unexplained problem in the light of such an interpretation.

I would suspect that an analysis along the line of Tolstoy's might demonstrate occupation at Portesuelo during the entire Classic period. This would then be a direct parallel to Azcapotzalco. It is certain that Brainerd's interpretation was *not* based on Tolstoy's evidence for a full period occupation at Azcapotzalco.

TOLSTOY'S SURVEY

The primary contribution made by Tolstoy is the carrying out of a comprehensive surface-collecting project at Valley post-Formative sites. The materials gathered by this activity have been utilized in making a sophisticated and carefully reasoned ceramic chronology by use of the seriation method so well established by J. A. Ford and his associates. Nearly two-thirds of the manuscript is devoted to technical type descriptions and periodization based on the seriation charts. Emphasizing the total ceramic complex as this method does, a new dimension has been added to Valley of Mexico archaeological studies. On the basis of Tolstoy's work it is now possible to make use of not only the numerous monochrome and plain sherd materials, but also to work realistically with some of the extremely rare varieties, including the very distinctive decorated types. We have thus, a fairly complete series of cross-sections of post-Formative ceramics from a number of sites.

In order to help align these surface collections into chronological series, Tolstoy utilized both unpublished data from Vaillant and published data from Acosta and Noguera. He also had a partial view of the preliminary results of our dig at El Risco. Although it is not elaborated in his report, at an early stage in analysis Tolstoy considered seriously that his evidence showed a transition from Classic to Post-Classic periods. The evidence for this was mostly from the Azcapotzalco zone. Further study, however, caused him to reject this idea and instead to conclude very strongly that he had no substantial evidence for such a transition.

Perhaps the most major point Tolstoy makes is for this sharp break or gap between Classic and Post-Classic periods, as indicated by his evidence of surface materials. And it is just this

point which I feel is controverted by evidence from El Risco.

Classic samples from the Azcapotzalco zone are quite different from those found in the Teotihuacan zone. Supporting Sejourné (1954), Tolstoy explains these differences as "regional" ones and equates the full span of Vaillant's unpublished data from El Corral with the Classic (II, III, IV) Teotihuacan sequence.

On the basis of his interpretation that there is no Classic to Post-Classic transition, Tolstoy sets up two mutually exclusive series of pottery types (San Martin for Classic and Tula or Texcoco for Post-Classic). On the basis of his interpretation of Azcapotzalco materials Tolstoy sets up two varieties of Classic culture (Amantla from Azcapotzalco and Teotihuacan from San Juan Teotihuacan). Most of the comparisons to be made with El Risco revolve about these two major points. There are, however, a number of other points to be discussed below.

General El Risco Comparisons

El Risco evidence seems to me to point directly to a transition, or rather a continuity between Classic and Post-Classic. On internal evidence alone we feel that the earliest period at El Risco is well established. The fact that this period can be compared closely to the El Corral II (Tolstoy, 1958: 60) phase at Amantla suggests naming it "Classic." It is impossible to make one-to-one comparisons here because the Amantla materials are typed into a different series from that of the materials from most of the cut at El Risco. As listed above in the periodization, however, the several distinctive San Martin and San Miguel types as well as the generally parallel trends in monochrome types suggest that the Classic period at El Risco is closest to El Corral II at Amantla. Most significant of all, it seems to me, are the trends in rim form and vessel form. At El Risco there is very definitely a shift in the lower levels in the direction of the Amantla Classic form congeries suggested by Tolstoy.

While our theoretical and typological approach is very close to that detailed by Tolstoy it is difficult (if not impossible) to demonstrate a transition within this framework. The basis for series separation is largely on Tolstoy's suggestive evidence for striking differences in vessel form and rim form among sherds otherwise indistinguishable. Neither his data nor ours are quantitatively adequate to prove this point, but I feel that the generally larger samples at El Risco lend support to the significance of our trends. The use of Vaillant's data lends considerable weight to Tolstoy's seriation study, but Vaillant's types are so divergent in both concept and typology that specific comparisons are difficult.

Notwithstanding the difficulties of establishing a Classic to Post-Classic transition, we feel that the El Risco evidence is very suggestive of such a stage of development. There are also a number of significant features of difference with Amantla that are apparent at El Risco. While it would be possible to suggest here another "regional" culture variety and let it go at that, the nature of the El Risco site (a simple rural unit) implies something else. I feel there is reasonable evidence suggesting a working hypothesis to explain both the Amantla-Teotihuacan and the Amantla-El Risco differences. The basis for these differences is the social structure and concomitant settlement pattern variations which are involved in the organization of Classic society.

SITE INTERPRETATION

SOCIAL STRUCTURE HYPOTHESIS

Data for the Valley of Mexico Classic period are essentially limited to three major archaeological zones (Teotihuacan, Portesuelo, and Azcapotzalco) which represent two significantly different kinds of settlement complexes.

Teotihuacan is the religious and ceremonial center (Vaillant, 1944), dominated by monumental architecture. The residences known to be associated with this center are quite specialized and apparently were used for and by only certain classes (the religious, priestly, and perhaps political leaders) of the Classic period society. This is reflected in the quality of the buildings, their decoration, and the kind of remains found as refuse and grave goods here. It also seems reasonable that this should be the case in terms of current developmental theories (Willey and Phillips, 1957) as well as Thompson's (1954) characterization of Classic culture. Additional support for this interpretation comes from the absence of certain items at Teotihuacan. I refer to the simple domestic and utilitarian objects (metates, molcajetes, comales, simple ceramics) and simple domestic architecture. These items are not well known (if at all) from Teotihuacan but they are quite important at El

Corral, Coyotlatelco, Portesuelo, and El Risco.

The Amantla area is much less well known during the Classic period but what evidence is available suggests a very different settlement complex from that of Teotihuacan. Vaillant's interpretation of the zone as a different kind of "Toltec" (i.e., Classic) settlement is based upon both historical and excavation information. Tolstoy and Sejourné each depend upon specific kinds of ceramic evidence for suggesting that the Amantla representative of Classic culture is a regional variant of Teotihuacan. Tolstoy, at least, would reject the idea of a Teotihuacan V period represented only at Azcapotzalco. He sees the ceramic evidence from his survey and Vaillant's El Corral excavations as paralleling the Teotihuacan II, III, and IV type site sequence in time, but differing from it in a number of specific material trait details. While there is some evidence for small ceremonial structures in the Amantla area (Gamio, 1921), most building activity was apparently in terms of domestic or residence complexes (Tozzer, 1921, and Vaillant, 1944). Most of the specific trait differences with Teotihuacan are in figurines and in simple and/or utilitarian artifacts, e.g., comales. (This interpretation of the Azcapotzalco zone is similar to the one advanced above for the Portesuelo zone.)

Faced with two parallel Classic sequences of related but differing sets of material it seems to me that describing them as "regional variants" is but an initial conclusion and begs the real question of relationship. Tolstoy's basic seriation study is a firm foundation for more searching explanations. In the light of inferential knowledge of Classic social structure why not hypothesize a population distribution in accord with this structure and explain differences in material remains as a correlate of this? Thus, we could understand the absence of much utilitarian material and domestic architecture at the religious and ceremonial center. For here mainly the privileged classes (the religious and intellectual hierarchy) lived and their everyday equipment and especially burial furniture were quite distinctive from that used by the middle and lower class groups. By the same token, the Amantla area and the Portesuelo zone take on real meaning if considered as commercial and middle class (artisans and merchants) residence and working zones. Ceremonial and religious associations here are on a lower level—local shrines and graves of the class; and there should be evidence for local manufacturing as well as ordinary domestic activities.

In the light of the size of the Teotihuacan zone and the lack of any other comparable Classic center, we agree with Vaillant's hypothesis that "Teotihuacan was probably the result of the cooperation of communities scattered over a large part of the Valley. Its scale and vastness could not have been achieved by a single resident community" (Vaillant, 1944: 158). Such a theory is a logical interpretation of the present evidence and as such it dovetails with the inferred Classic social structure and population distribution.

While there apparently is (as Vaillant, 1944, suggests) a completely different situation in Post-Classic times with regard to scale of activities and demands upon population, the kind of settlement represented at El Risco is at the simplest end of the folk-urban continuum. The simple rural homestead or hamlet is perhaps the longest lived population unit in the Valley, being present from earliest Formative times up to the present, and numerically most important during Formative and Classic times.

In summary then, we would explain the differences among Teotihuacan, Amantla-Portesuelo and El Risco late Classic units as a reflection of three major kinds of settlement based on three main strata of Classic society:

Community	Settlement type	Social units	Archaeological units
Urban	Ceremonial center	Religious and intellectual hierarchy	Teotihuacan
Peasant	Secular center	Artists, craftsmen, merchants, traders	Amantla and Portesuelo
Folk	Rural loci	Farmer, laborer	El Risco plus others

This interpretation of the present evidence amounts to an hypothesis for explaining cultural differences among Valley of Mexico units thought to be contemporaneous. I have suggested, in effect, using the concept of a folk-urban continuum in a spatial rather than a temporal sense, and this as a simplistic description of what is most certainly a more complex social structure and residence pattern. In Classic times there certainly were rural, folk units as the basic population sub-stratum. Almost as assuredly there were peasant groups of full-time and part-time specialists (i.e., non-food producers) to provide

the skilled labor and creativeness needed for production of the high quality, monumental and luxury items consumed by the highest class levels. This priestly hierarchy we equate with an urban elite, using the term urban in a sense somewhat removed from the conception of urban in Post-Classic times. Since we would assume growth of urbanization trends out of the folk and peasant units, the Classic expression of urbanism is developmental rather than final.

Specific El Risco Comparisons

Having essentially accepted the several assertions (Vaillant, Sejourné, Tolstoy) indicating the divergent (from Teotihuacan) Classic nature of Amantla deposits, I have used this fact of cultural variety to suggest the above interpretation of urban and peasant settlements. On purely theoretical grounds we would posit the ultimate folk or rural unit as the third major kind of settlement to be expected in the Classic stage. We would expect it to be most remote culturally from the urban elite assemblage. As a corollary we would expect it to be most similar to the peasant or secular center. We would expect fairly direct parallels with the secular center since most direct and frequent contacts and communication would be between these two units rather than between rural and urban units. The secular center functioned as the intermediary and thus should be the connecting link for typological and temporal equations of folk and urban units.

This theoretical rural locus is apparently the sort of unit we have at El Risco. The entire site is a complex of chinampas, presumably with associated single family adobe houses, strung out along the western margin of Lake Texcoco about five miles northeast of the Azcapotzalco zone. No examples of complex multi-roomed domestic or ceremonial structures are yet known from the immediate site area. The question of midden growth here is a basic problem. Were these units built by unconscious growth and accretion under day-to-day conditions of pot breakage and dispersal? Or were they specially built in periodic increments of sherd dumping? Because of the relative smoothness of type trends in Cut A, a gradual "normal accretion" growth is indicated. The impressively high number of sherds and the proportion of sherd to dirt material on the other hand combine with the lack of bone and stone refuse to suggest intentional selection and dumping of sherd refuse. The latter possibility is also suggested by historical data on intentional building of chinampas as garden plots or other extensions of land surface.

While these questions about the specific nature of this rural domestic unit are substantive we are apparently not yet possessed of definitive evidence one way or the other. A major question to be raised is in the nature of our El Risco evidence for the suggested transition between the Classic and Post-Classic periods. Although I do not feel that a detailed knowledge of a late Classic rural unit is essential to formulating the above structural-demographic hypothesis it is essential to substantiate the theory on all levels.

Because we have sampled only one of many tlateles present at the El Risco site our interpretation of this one location may well not hold for all other units at the site. The nature of the site as a whole, I suggest, is an important problem to be investigated. The observations we have made, however, suggest that most of the tlateles are very similar in construction, size, and presumably contents.

SALT-MAKING THEORY

Tolstoy has examined the possibilities for considering Texcoco Fabric-marked as a pottery type specialized in function for salt-making purposes. He suggests explaining the temporal and geographic distribution of this type in terms of a possible source (the incoming Tenochcas) and urban expansion of the fourteenth and fifteenth centuries. Seen from the solitary vantage point of El Risco, the length of time during which Texcoco Fabric-marked is prominent suggests (accepting the functional explanation) the relative permanence of a location for the industry. Having searched for sites producing this type it is significant that we returned to El Risco for our excavations. Here is found the most extensive and complex site producing the type. While factors of salinity were no doubt paramount in determining the original location for the early stages of a salt-making industry, establishment of a local specialization perhaps tended to perpetuate and concentrate the activity.

On the basis of Sanders' (1956) descriptions of present-day regional specializations, we suggest that El Risco is a pre-Spanish parallel; it early became an area of salt production and manufacturing. Its product was perhaps distributed in the locally-made package (the Texcoco Fabric-

marked pot) and thus distribution of the type is not to be understood only in terms of local ceramic manufacture and local salt production.

No doubt the processing was done in other suitable locations, but this "industrial zone" interpretation of El Risco makes sense in the light of our broader interpretation of the various Classic settlement units. Perhaps El Risco represents an area of a developing "cottage" industry. The similarity of shape for many of the early Texcoco Brown vessels may even indicate that Texcoco Fabric-marked was not the earliest vessel specialized for this manufacturing process.

POPULATION INFERENCES

In the context of our assumptions about the nature of Classic and Post-Classic societies (*cf.* Willey and Phillips, 1957) we suggest that each of these tlateles represents the accumulation of debris from a small social unit. These are not "village" sites, but rather are perhaps extended family dwelling and working areas. It is possible that occupation was not completely permanent since the cultivatable land area was no doubt quite restricted at times of high lake level. These tlateles may represent involuntary islands resulting from salt-making activities, but functioning also as house platforms and small milpas. The population of the area, in this view, would thus consist of scattered and separate households (from 50–150 yards apart) clustered along the shores of a saline bay about two miles long. Specific studies of a representative sample of all El Risco tlateles could elucidate this matter, I am sure.

Considering the site at one time level, like this, we see a possible structure for the community. But we know that occupation in cut A carried over a period of at least 600 years. There is ample opportunity for many changes in the community within such a span of time. Presumably certain parts of the site were occupied earlier than others, probably even earlier than the base of cut A. Certainly we would expect that salt was an important necessity from very early times. The early Classic period developments of trading and specialized manufacture imply that a basic commodity like salt would have been early in any list of such manufacturing developments. The presence of F_2 rims of Texcoco Brown in the Toltec and Teotihuacan levels of cut A indicate a vessel form like the Fabric-marked (even to a roughened exterior wall) which may have functioned as a salt-evaporating vessel prior to the acceptance of the Fabric-marked style. The El Risco location is on the margin of what may have been one of the most saline portions of the lake. This suggests that, originally, site location was not determined primarily by factors important to agricultural production. The numbers of people at El Risco and their distribution in residences appear then as a probable function of a role in the larger community (i.e., a manufacturing district) during both Classic and Post-Classic times. If, by testing this hypothesis we could establish the presence of a salt-making industry at El Risco in Classic times, this fact would be a conclusive indication of lower class level continuities between the two periods.

One is hard put to see salt-making techniques and, more importantly, the need for salt as being the exclusive concern of Post-Classic society. The breaks in the sherd record are strong indicators for basic period differences, *but* different clusters may characterize the temporal distribution of traits other than sherds. Any final or cultural periodization should be based on several kinds of evidence and should not ignore the evidence for continuity.

An interesting point seen from the pottery type distributions is the parallelism between the multiple frequency peaks of Texcoco Orange and both Tenayuca Black/Orange and Tenochtitlan Black/Orange. While this is not conclusive evidence for the ceremonial destruction of ceramics in observation (see Nicholson, 1955) of "New Fire" ceremonies, it does suggest a cyclical rather than a simple unimodal graphic expression of ceramic popularity. Dealing with a complex, urban civilization as we are one can question, on this basis, the exclusive application of the simplistic unimodal distribution theory of culture change. This seems a more likely explanation than population level fluctuations. If so, however, there must be some long time lags represented in the Post-Classic periods at El Risco.

SETTLEMENT PATTERN INFERENCES

Tolstoy's interpretations follow traditional lines in seriation work—differential spatial and temporal distributions are considered to be data implying unspecific "cultural influences" from this area or that at one time or another. While such implications are certainly significant, I feel that greater emphasis on "internal" explanations

or specific kinds of cultural influence should shed alternative and supplementary light on the complex pattern of civilizational growth and development.

While the Tepeyac area (around El Risco) is considered by Tolstoy to be most closely related to the Atzcapotzalco Classic materials on typological grounds, it is also considered most "backward," i.e., late in achieving various Post-Classic ceramic developments. Tolstoy suggests consideration of industrial or ecological factors as well as a possible "ethnic" barrier here. This inference is quite interesting when viewed in the light of Sanders' statements (1956: 120) about the divergent modern settlement patterns of Otomi-speaking groups in the northern Valley. Could we have at El Risco a pre-Spanish parallel for the modern situation? Does the modern but marginal (settlement-wise) group of Otomi represent a relic from former times? Was Classic culture Otomi-speaking? Such questions spring to mind but closer to our present evidence we can suggest some ideas on settlement and residence factors that are still far from clear.

In his paper on Valley of Mexico settlement patterns, Sanders (1956) discusses both "community" and "zonal" patterns. Under the former he distinguishes: the contemporary units of village, town, and city; the comparable ethnohistoric units; and the archaeological zone at Teotihuacan with its urban center, residential barrios and workshops (on this we have very little basic field information but Sanders stresses the fully urban nature of Teotihuacan). The zonal patterns are discussed essentially from only the contemporary point of view. Stress here is broadly ecological —a zonal settlement pattern is the expression of a functional symbiosis based upon natural environment, various adaptations to this and commercial-political interrelationships based on these two factors. While I am not entirely clear what Sanders means by zonal settlement patterns, I think we can apply our version of his concept to the pre-Spanish period more explicitly than he does.

This concept is more meaningful to me if we distinguish first of all the scope of the community concerned, then look at it alternately in morphological and functional terms.

Classic Settlement Patterns

Our social structure hypothesis (elaborated above) suggests that Classic society comprised a single broad community in the Valley of Mexico. In a very real sense (tribute, work contribution, and religious dependence upon one major ceremonial center) this community was different from the Post-Classic "city-state" units as well as the unifying Toltec and Aztec political domains. The scope of our community in Classic times then is largely delimited by the natural area—the Valley. Within this were certainly smaller and more intimately interrelated communities.

Morphologically we can see the Valley in Classic times as comprising one major architectural center (Teotihuacan), two major residence centers (Atzcapotzalco and Portesuelo), and a number of rural loci, one of which is El Risco.

Functionally, the zone of major architectural activity can be described as an urban ceremonial center and upper class residential zone. The two residence zones are major urban loci of commerce, manufacturing, and other secular activities, as well as residential zones for the middle classes involved in these activities. The rural locations to be expected are many with a variety of possible symbiotic specializations. Chinampa farming and lakeshore fishing communities probably existed. There no doubt were others. At El Risco we may have an early and basic manufacturing unit—salt production—and the community here was perhaps also food producing, but may have been dependent upon other rural units for much of its food supply.

Post-Classic Settlement Patterns

The Post-Classic patterns appear to differ markedly from those of the Classic. There is apparently a breakdown of the Valley-wide sense of community with the fall of Teotihuacan and the rise of Tula. Toltec ideas and material goods supplant Classic units in the Valley. From the historic records of Aztec times the picture of relatively independent city states emerges as characterizing Post-Classic times. Aztec communities, perhaps starting with Texcoco and Tenayuca and continuing with the rise of Tlatelolco and Tenochtitlan, are characteristically smaller in scope than those of the Classic and are related to each other in more formal and secular as well as political ways. The basic sense of community, however, is more localized.

Morphologically, the Post-Classic can be viewed as a time of independent city-state developments, with two major overall unifications. First by

Toltecs from Tula, and later by the Tenochcas from Tenochtitlan. These city state units were probably the most clearly urban in our modern sense of the term. Communities such as Tenochtitlan included both religious, commercial, and residential zones within their boundaries as well as fringe or suburban farming zones. This is seen as a quite different pattern from our postulated Classic morphology.

Functionally, the rural units including various specializations, e.g., salt-making, apparently continue with little change from the Classic pattern; except, of course, where physical expansion, as of Tenochtitlan, infringes and absorbs. The secular units and ceremonial unit of Classic society zonal pattern are now included in each of the large urban city-state zones, but community pattern of residence, commerce, and religion are still reflected here in spatial distribution within the city-state.

In Classic times the zonal pattern is Valley-wide and community pattern is closely integrated with it. In Post-Classic times there are probably several zonal patterns which may be differentiated from the several community patterns, though functionally interrelated with them.

The above settlement-pattern inferences are highly speculative and as yet do not form a consistent hypothesis. They do imply a basic difference between Classic and Post-Classic. In the earlier period a symbiotic relationship existed integrating all settlements in the Valley on a religious and ceremonial basis. This broke down for reasons largely unknown with local units becoming independent in the Post-Classic until united on a military and political basis by Tula and later Tenochtitlan. We see, then, in this view, a new dimension of complexity to settlement pattern studies, especially in the Post-Classic period.

COMPARISONS

Four papers in the Viking Fund settlement pattern volume (Willey, 1956) dealt with Meso-American data. We have already discussed some points presented by Sanders (1956). Shook and Proskouriakoff (1956) describe the highland Maya development in the Guatemala valley as follows:

> In summary, we can say that there were three successive patterns of settlement in the Guatemala valley: first the Pre-Classic pattern, characterized by a distribution of sites over most of the arable land, with one very large, centrally located town, and by a typical open assemblage of large ceremonial buildings; second, the Classic pattern, with a limited site distribution in the valley and with more closely knit ceremonial centers, which in Late Classic times almost invariably include at least one ball court; and, third and last, the pattern of conquest times, of only small settlements in the valley and of towns strategically located on high ground and protected by outposts (p. 99).

In this same volume Borhegyi (1956*b*) and Willey (1956*b*) present contrasting points of view on highland and lowland Maya developments. Borhegyi suggests a strong correlation between contemporary and pre-Columbian patterns as paraphrased below.

Borhegyi sees the plaza as the important focal point of activities unifying present-day Guatemalan Indian communities and suggests allying these modern units with comparable pre-Spanish communities.

Post-Columbian	Pre-Columbian
A. Dispersed villages	Scattered or agglutinated small-house village (with no plaza)
B. Compound villages	Compound or agglutinated hill-top village (with plaza)
C. Concourse centers	Ceremonial or pilgrimage center (plaza extremely important)

In another article Borhegyi (1956a) has stressed the apparent independent and parallel development of a folk level in highland Maya society. He does not attempt to relate this folk unit with any specific settlement units, although the scattered small house village seems a likely correlate.

Willey (1956*b*), however, has provided three ideal types for lowland Maya settlement pattern and has suggested or rather implied correlations with the level of folk-urban complexity.

In contrast to Borhegyi's views on the Highland, Willey (1956*a*) sees the lowland folk units in the society as participating to a fair degree in the knowledge and values of the more urban classes. This is reflected in the settlement patterns which include ceremonial structures of a lower order than those in the main centers.

By comparing these two different Maya settlement patterns with our suggested Classic pattern for the Valley of Mexico several points of both similarity and difference are brought out. The highland Maya pattern of scattered or agglutinated small-house villages can perhaps be equated with pre-plaza outer portions of Willey's type "B," and our Classic rural loci. The highland compound village with plaza is apparently iden-

tical with Willey's type "B" and seems to compare closely to our Classic secular unit. The highland ceremonial centers are a straight equation with Willey's type "C" and our Classic ceremonial center. Here we are comparing essentially the community patterns, unrelated to a zonal scheme for the Maya area, but parts of a specific Classic zonal pattern in the Valley of Mexico.

If we compare these same Maya patterns with our Post-Classic patterns we see how really inadequate any of these approaches are. In Willey's scheme there is no real provision for the highland scattered village type which compares with our rural loci that continue essentially unchanged from Classic to Post-Classic. The compound villages and ceremonial centers of the Maya highlands (and Willey's types "B" and "C") are essentially amalgamated in our scheme for the urban city state in the Valley of Mexico.

We are not really searching for an overall scheme that will include and correlate a number of disparate Meso-American developments. But this attempted comparison points up the need for much more rigorous control of a number of important factors—first of all the level or scope of the "community" under consideration (a clear distinction between zonal and community patterns), then the distinctions between morphological and functional factors distinguishing community units, and finally, a clear understanding of historical and developmental elements.

Although in our discussions of site inferences above we attempted to restrict our settlement pattern inquiry to structural and functional considerations, we cannot completely ignore the developmental aspects of settlement problems. The recent seminar report (Meggers, in Wauchope, 1956) has suggested a series of developmental stages knowledge of which has been implicit in many of our statements. There is definitely a time depth involved in our study of El Risco, but we have suggested little or no change in the nature of this settlement. Growth in size, yes, but functional or developmental change, no. At other levels of Classic and Post-Classic societies there are factors which support the seminar group's allocation of Teotihuacan (Classic society) to the advanced nuclear centered stage and Tenochtitlan to the supra-nuclear integrated stage. (Why Tula, in the light of expansion to Yucatan, is not included here is puzzling.) In the light of our suggestion for Classic period secular centers perhaps Classic society really represents the supra-nuclear stage of integration.

While the typological units derived by both Borhegyi and Willey are essentially static in concept, the scheme proposed by Shook and Proskouriakoff is essentially developmental and concerned with a maximum-scope community. (In the light of our continued lack of knowledge of pre-Formative units in Meso-America, the priority in time and developmental stage for highland Guatemala invites suggestions for trans-Pacific influences.) The Guatemala Pre-Classic pattern could include all of Borhegyi's structural units and compares closely with the Classic zonal pattern as we have suggested it for the Valley of Mexico. The Guatemala Classic pattern is a different arrangement of Borhegyi's units, emphasizing the ceremonial centers. There are parallels here to the Valley of Mexico Post-Classic pattern but important differences on the level of political and military factors. The conquest pattern in Guatemala is another arrangement of Borhegyi's units but apparently has few parallels in the Valley of Mexico.

FUTURE RESEARCH

As a result of our work at El Risco a number of questions and problems for future research have come to light. Most of these we have touched on in some way, a few we will try to follow up in our own future Mexican work. Others are simply suggestions of important problem areas.

A. An intensive study of the salt-making industry in the Valley should be a profitable field of intensive effort. Studies of the El Risco site, the Ixtapalapa site and the distribution of Texcoco Fabric-marked are important leads here. The bringing to bear of ethnohistorical sources and a broad comparative approach promise fruitful results.

B. A number of specific archaeological sites or zones should be studied intensively as occupation areas for elucidation of community settlement patterns. Teotihuacan, Portesuelo, and Azcapotzalco (though now much built over) should help clarify the Classic period. Tula, El Risco, and other rural loci should provide Post-Classic data.

C. In conjunction with B above, aerial photo surveys of especially the Teotihuacan, Portesuelo, and Tula zones should be instructive of settlement pattern data. This approach may

also be useful in a study of tlateles or chinampa communities in general.

D. There is need for more intensive stratigraphic excavation work in both Classic and Post-Classic period sites to help adjust the confusion presently apparent in seriation ceramic typology between the two periods. Significant breaks in distribution can apparently be found more readily in categories of form rather than simple slip color.

E. Specific search should be made for Classic communities in the Valley.

F. A specific search for archaeological remains of specialized communities such as fishing and manufacturing units may help elucidate aspects of the zonal settlement patterns.

G. A specific study of changes in the natural environment and the possible correlation of these changes with cultural factors. (Lake level changes seem particularly important here.)

H. An examination of the zonal and community pattern concepts should be made in detail in both theoretical and comparative terms.

I. A study of the developing "urbanism" of the Valley of Mexico in terms of the suggested contrasts between Classic and Post-Classic varieties of the urban quality.

J. A detailed ethnohistorical approach to Aztec social structure with an attempt to correlate this with Aztec period archaeological remains and extend it back in time.

ACKNOWLEDGMENTS

The basic project of field work could not have been accomplished without the financial support of seven Pittsburgh donors: P. W. Aitkenhead; Allegheny chapter, Society for Pennsylvania Archaeology; A. Bindley; E. Crump, Jr.; R. A. Fisher; Mitchell and Richey; E. O'Neil II. The encouragement I received from Carnegie Museum Director, M. Graham Netting, and Assistant Director, James L. Swauger, was made concrete in numerous helpful ways. The laboratory study and report writing have received financial support from the Carnegie Museum, the American Philosophical Society, and the University of Toronto, for which I am deeply grateful.

Various individuals have given specific help to various stages of the project. My wife, Nita, originally co-author of the projected report, not only fully shared the field work, but initiated the detailed sherd studies and has maintained a strong interest in the continuing analysis and write-up. Mrs. G. H. Lester and Miss Marcia Getting, both of Pittsburgh, volunteered many hours of tedious sherd-sorting work. John E. Guilday, Carnegie Museum osteologist, kindly identified the faunal material from the excavation. Both Clifford J. Morrow of Carnegie Museum (maps 1 and 2; figs. 1, 2, 5 and 8) and Zenon S. Pohorecky of the University of Toronto (figs. 6, 7, 8, 9 and 10) contributed valuable help by preparing the various illustrations listed. Peter Harrison and Eileen Bell, both of the University of Toronto, were helpful; Harrison made duplicates of selected illustrations and Miss Bell typed the complex table 9. Peter Kunkel of the University of California at Los Angeles was most helpful to me during my brief study of the Portesuelo collection. To all of these people I am greatly indebted and wish herein to express my thanks.

Dr. Paul Tolstoy has provided me with stimulating discussions on several occasions and very kindly sent me both an early and a final copy of his extensive manuscript as well as specific sherd type samples. Although our collaboration was very loose, we kept in touch and I think profited by this procedure. Tolstoy also read this paper in manuscript and offered comments. Other colleagues who have read and commented on the manuscript are T. F. McIlwraith, R. C. Dailey, R. W. Dunning, and W. T. Sanders. To each of these men I express my thanks for their helpful criticism.

In Mexico our project was encouraged and supported by Professor Eduardo Noguera, former director of pre-Hispanic monuments. He not only provided us with the specific legal permits needed for field work, but expedited a number of technical problems. I am particularly indebted to Sr. Rafael Orellana who helped me "push" the sherd collection through customs red tape in one day—in order to leave Mexico with the complete collection of excavated materials in our vehicle.

REFERENCES

APENES, O. 1943. The "Tlateles" of Lake Texcoco. *Amer. Antiquity* **9** (1): 29–32. Menasha.

DE BORHEGYI, S. F. 1956*a*. The development of folk and complex cultures in the southern Maya area. *Amer. Antiquity* **21** (4): 343–357. Salt Lake City.

——. 1956*b*. Settlement patterns in the Guatemalan highlands: past and present. *Viking Fund Publication No. 23*: 101–107. New York.

BRAINERD, G. W. 1951. The place of chronological ordering in archaeological analysis. *Amer. Antiquity* **16** (4): 301–313. Salt Lake City.

GAMIO, M. 1922. La poblacion del valle de Teotihuacan. 1. Mexico.

GRIFFIN, J. B., and A. D. KRIEGER. 1947. Notes on some ceramic techniques and intrusions in central Mexico. *Amer. Antiquity* **12** (3): 156–158. Menasha.

HICKS, F. MS. The chronological position and cultural significance of Coyotlatelco pottery of the Valley of Mexico. Department of Anthropology and Sociology, Univ. of California, Los Angeles.

LINNÉ, S. 1934. Archaeological researches at Teotihuacan, Mexico. *Ethnological Museum of Sweden, New Series, Publication No. 1.* Stockholm.

NICHOLSON, H. B. 1955. Native historical traditions of Nuclear America and the problem of their archaeological correlation. *Amer. Anthropologist* **57** (3): 594–613. Menasha.

SANDERS, W. T. MS. The urban revolution in central Mexico. Deposited in Peabody Museum library, Harvard University.

——. MS. Tierra y agua: a study of the ecological factors in the development of Meso-American civilization. Harvard Univ. Ph.D. thesis.

——. 1956. The central Mexican symbiotic region: a study in prehistoric settlement patterns. *Viking Fund Publication No. 23*: 115–128. New York.

SEJOURNÉ, L. 1954. Informe sobre el material exhumado en Ahuitzotla. VI, Mesa Redonda de la Sociedad Mexicana de Antropología—Mimeographed reports, V–I–1b.

SHOOK, E. M., and T. PROSKOURIAKOFF. 1956. Settlement patterns in Meso-America and the sequence in the Guatemalan highlands. *Viking Fund Publication No. 23*: 93–101. New York.

THOMPSON, J. E. S. 1954. The rise and fall of Maya civilization. Norman, Univ. of Oklahoma Press.

TOLSTOY, P. 1958. Surface survey of the northern valley of Mexico: the classic and post-classic periods. *Trans. Amer. Philos. Soc.* **48** (5). Philadelphia.

TOZZER, A. M. 1921. Excavation of a site at Santiago Ahuitzotla, D.F., Mexico. *Bull. Bur. Amer. Ethnol.* **74**. Washington.

VAILLANT, G. C. 1944. Aztecs of Mexico. Doubleday, Doran & Co., New York.

WAUCHOPE, R. (editor). 1956. Seminars in archaeology: 1955. *Memoir* no. 11, *Society for American Archaeology*. Salt Lake City.

WILLEY, G. R. 1956*a*. The structure of ancient Maya society. *Amer. Anthropologist* **58** (5): 777–783. Menasha.

——. 1956*b*. Problems concerning prehistoric settlement patterns in the Maya lowlands. *Viking Fund Publication No. 23*: 107–115. New York.

—— (editor). 1956*c*. Prehistoric settlement patterns in the New World. *Viking Fund Publication No. 23*. New York.

APPENDIX 1

Osteological Report—El Risco II, cut A

By John E. Guilday

Level I–N . . . . Dog, *Canis familiaris*, 1 cervical vertebra

Level VII . . . . Rabbit, genus *Sylvilagus* ?, 1 juvenile femur
Bird, species ?, large, 1 fragment of innominate
Unidentified, 2 fragments

Level VIII . . . . Dog, *Canis familiaris*, fragments of 1 femur, 1 ulna, 1 astragulus
Pocket Gopher, *Cratogeomys merriami cf.* 1 mandible, 1 incisor, 1 femur (from level VII)

Level IX . . . . Dog, *Canis familiaris*, 1 cervical, 1 thoracic vertebra, 1 calcaneum, 1 auditory bulla, 1 parietal (?)
Man, 1 phalange
Unidentified, 3

Level X . . . . Dog, *Canis familiaris*, 1 ulna fragment
Musk Turtle, genus *Kinosternon*, 4 fragments of plastron
Man, 38 fragments, adult
Unidentified, 6 fragments

Level XI . . . . Dog, *Canis familiaris*, 1 molar, 1 penis bone, 2 cervical vertebrae
Pocket Gopher, *Cratogeomys cf. merriami*, 1 mandible
Turtle, *Kinosternon* ?, 1 fragment of carapace
Man Adult—1 cuboid, 1 fragment of zygoma, 1 distal end of humerus, 1 phalange—(big toe)
Infant—2 femora, 1 humerus, 1 ulna, 1 ilium, 1 tibia, 1 temporal, 1 additional skull fragment
Bird sp. ?, 1 fragment of radius

Level XII . . . . Dog, *Canis familiaris*, 1 femur, 1 innominate, 1 tibial epiphysis (at least two dogs here)

Minimum number of individuals: Dog 4, rabbit 1, pocket gopher 2, mud turtle 1, bird 1, man 2.

Kinosternon is a small, inedible, aquatic turtle.

Cratogeomys, about the size of a large rat, a serious agricultural pest and could conceivably have been eaten although the bones bear no witness on this point.

Canis 1 size of terrier, 1 size of large spaniel, actual breed ?.

ROMAN MEN IN GREEK HIMATION (ROMANI PALLIATI)
A CONTRIBUTION TO THE HISTORY OF COPYING *

MARGARETE BIEBER

INTRODUCTION

GREEK art in the classical period, from about 500 to 330 B.C., aspired toward ideal truth as formulated by Plato. Eternal ideas of beauty, goodness, truth, and justice were embodied in literary and artistic creations. At the same time, artists also began to imitate nature, or rather attempted to represent according to the laws of nature. This led to naturalism and greater emphasis upon real life and, especially, upon the individual. This development took place from the fourth to the second centuries B.C. Then, in the second century of the Hellenistic period began the adaptation of classical masterpieces to contemporary tastes and needs. Copying flourished for the use of the Romans during their late republican period of the first century B.C. Then the tradition of copying was carried on through the whole imperial period.

We might even look upon modern attempts to imitate ancient forms as a fourth degree of copying (see figs. 2*b* and 51).

Since, to a large extent, we depend on copies of masterpieces for our knowledge of Greek art, many scholars have worked on problems associated with such copies. How can a Greek original be distinguished from a later copy? Which are the Greek and which are the Roman elements of the copy? Which are the best copies to use in order to understand the earlier art?

Most scholars have worked with nude male statues. But there is always the possibility that the copyist has studied the anatomy and facial expressions of his contemporaries. For example, it is still not possible to ascertain whether the harsher or softer copies of the Diadumenos of Polykleitos approximate the original. When a statue contains a portrait head, it is easy to extract the Roman elements and date the copy. Thus the so-called Germanicus of Kleomenes in Paris was copied from a statue of the early classical period. But it has a portrait head of the Julio-Claudian period, perhaps of the young Octavianus. The Hadrian in the Capitoline Museum has the portrait of this emperor on a statue of the "Ares Borghese," which was created about 400 B.C. These statues can contribute to the history of copying. However, they are individual cases, and it is not possible to build up such a history from them alone.

I myself have studied particularly draped statues, and up to now mostly female statues. The draping gives subjective criteria for judging the copies, in so far as the copyist has often misunderstood the Greek dresses. They were sometimes no longer worn in his period, or they had such an artistic arrangement as the copyist was not able to understand. Mistakes made in the Roman period, when the dresses were mostly fitted and not freely draped, betray the inexact copyists.

GLOSSARY

Adaptation: Modification to make suitable for a purpose.
Aischines (Aeschines): Greek orator, partisan of Philip of Macedon, adversary of Demosthenes, *ca.* 390–316 B.C.
Arringatore: Ital. = orator. Name given to the statue of the Romanized Etruscan Aule Metilius by the artist Tenine Tuthines.
Chiton: Garment worn next to the skin.
Copy: Exact imitation of an original.
Curculio: Name of parasite in the comedy of the same name by Plautus.
Fabula Palliata: A Latin play imitating a Greek play and presented in the Greek cloak.
Heautontimorumenos: Greek. The Self-Tormentor. Comedy by Terence imitated from Greek comedy, presented in 163 B.C.
Hecyra: Greek. The Mother-in-Law. Comedy by Terence, imitated from Greek comedy, produced in 165 and 160 B.C.
Himation: Greek mantle. A garment consisting of a rectangular cloth draped freely around the body.
New Comedy: The comedy of the Hellenistic period, created by Menander at the end of the fourth century B.C.
Palla: Lat. mantle = the Greek himation, worn by women.
Pallium: Lat. mantle = the Greek himation, worn by men.
Plautus: Latin comic poet, *ca.* 254–184 B.C.

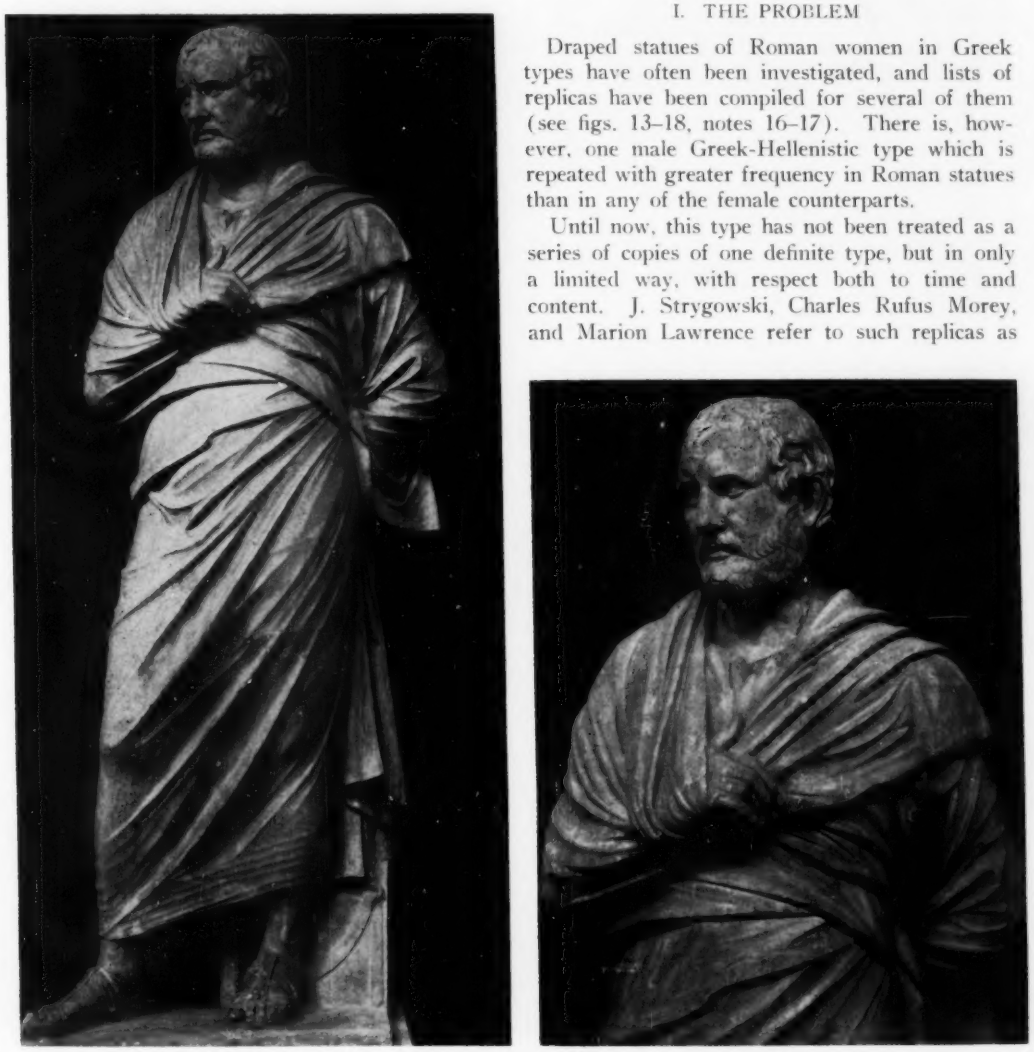
* In 1951–1952, with the aid of a grant from the American Philosophical Society, I traveled in Europe for the purpose of collecting material on Greek draped statues and Greek portraits made during the Roman period. See *Year Book Amer. Philos. Soc.* for 1952: 283–285, 1953.

Pudicitia: Lat. = modesty. Name given to a statue in a modest and restrained pose.
Quintillian: Latin orator, A.D. 37–95.
Replica: Duplicate, exact copy of a work of art.
Scrinium: Box for rolled manuscripts, book case.
Scroll: Rolled manuscript, the ancient form of books, written on rolls of parchment or paper.
Sophokles: Greek tragic poet, 496–406 B.C.
Sidamara: Place in Asia Minor where Roman sarcophagi of the second and third centuries A.D. have been found.
Sportula: Lat., basket.
Stele: Greek and Latin. An upright slab or pillar of stone used as a gravestone or for inscriptions.
Toga: Latin. Outer garment worn by the Romans when appearing in public. Large half-elliptical shape.
Tunic, Tunica: Lat. undergarment, corresponding to the Greek chiton.
Type—Lat. typus: Object or conception serving as model for subsequent works of art; characteristic specimen of a group.

I. THE PROBLEM

Draped statues of Roman women in Greek types have often been investigated, and lists of replicas have been compiled for several of them (see figs. 13–18, notes 16–17). There is, however, one male Greek-Hellenistic type which is repeated with greater frequency in Roman statues than in any of the female counterparts.

Until now, this type has not been treated as a series of copies of one definite type, but in only a limited way, with respect both to time and content. J. Strygowski, Charles Rufus Morey, and Marion Lawrence refer to such replicas as



a *b*

FIG. 1*a–b*. Statue of Aischines, Naples. Phot. Alinari, nos. 11070 and 34252.

FIG. 2*a*. Statue of a man in the draping of Aischines. Verona, Museo Civico, no. 145. Phot. German Arch. Institute Rome, no. 8302.

FIG. 2*b*. Model of draping of Aischines. Phot. Bruno Schroeder.

the "Sophokles" type.[1] The latter two have treated this type only in relation to the Sidamara sarcophagi, classifying it as No. 1 and listing No. 17 as a variation. Ernst Buschor has called it the Aischines type.[2] Friedrich Wilhelm Goethert and Robert West have named the mantle the

[1] Strygowski, P., *Orient oder Rom*, 58–59, fig. 22 (Sophokles). Morey, Charles Rufus, The sarcophagus of Claudia Antonia Sabina and the Asiatic sarcophagi, in *Sardis* **5**(1): 12, 37–48, 60, No. 1; 63, No. 17; 66. Frontispiece, figs. 9, 11, 25, 55, 105 (type 1) and 120 (type 17. Christ, with changed standing leg); fig. 134 (Sophokles), 1924. Lawrence, M., in *Memoirs of the American Academy in Rome* **20**: 143–144; *cp.* 134–135, fig. 19; 42, fig. 29; 144, fig. 31, 1951.

[2] Buschor, E., *Das Hellenistische Bildniss*, 8, München, 1949. For Aischines (fig. 1) see Arndt, *Griechische und römische Porträts*, pls. 116–120, 641–644. Brunn-Bruckmann, *Denkmäler*, pl. 428. Hekler, A., *Bildniskunst*, pl. 53. Laurenzi, L., *Ritratti Greci*, no. 46, pl. 17. Schefold, K., *Die Bildnisse der antiken Dichter, Redner und Denker*, 102–103. Wilson, Lillian M., *The clothing of the Ancient Romans*, 80–81, fig. 49, Johns Hopkins University Studies in Archaeology, No. 24, Baltimore, 1938. *Cf.* our model for the draping of Aischines (fig. 2*b*).

toga exigua.[3] Others, including West, preferred to see the men wrapped in the mantle as depicting philosophers or orators.

In the light of available evidence, these interpretations appear neither adequate nor comprehensive. Figures of men wearing the pallium, copied again and again by the Romans, did not represent the Sophokles or Aischines types. Nor was the mantle a toga exigua. Nor were all the men garbed in such clothes philosophers or orators. Nor were such artistic representations confined to sarcophagi beginning in the second century. Rather, this distinct art form had a long and continuous history which spanned more than six centuries. Transposed from Greek to Roman culture during the Hellenistic period, it persisted right up into the early Christian era.

II. THE GREEK MODEL

The statues of Sophokles and Aischines are distinguished by a proud and self-sufficient attitude. It is expressed particularly by the left hand placed on the hip and the right hand laid before the breast. Both statues were created about 340–330 B.C. They are thus art forms of the late classical period. The statue of Sophokles was not created during his lifetime. Rather it was erected in the theater of Lykurgus at Athens as an ideal conception of how Sophokles must have looked one hundred years earlier and how he lived on in the hearts of his countrymen. He is represented as the ideal of a man equally handsome and good (kalos kagathos καλὸς κ' ἀγαθός). In contrast, Aischines (fig. 1) is represented in a contemporary statue, probably made at the height of his career as a partisan of Philip and Alexander of Macedon. He strikes a pose which borders on conceit. While the right hand of Sophokles is laid loosely and relaxed on the right side of his chest, the right hand of Aischines hangs stiffly in the middle of his chest. The two statues have quite different characters. There is no repetition in classical creations. There are a few examples of statues in the Aischines type: the portrait of a young man in Kopenhagen, No. 462; a statue in the Museo Civico in Verona (fig. 2); the terracotta statuette of an older man in the National Museum at Athens (fig. 3); a torso from Samos

FIG. 3. Terracotta statuette of an older man. Athens National Museum. Phot. German Arch. Institute, Athens, N. M. 199.

belonging to the first century B.C.; and a torso from Prusias, now in Istanbul.[4] The Aischines

[3] Goethert, F. W., *Die Kunst der römischen Republik*, 15–22, Berlin, 1931. *Idem* in Pauly-Wissowa, *Real-encyclopädie*, ser. 2, **6**, *s.v.* toga, 1651–1660, 1937; West, R., *Römische Porträtplastik* **1**: 41–44, figs. 27–29, München, Bruckmann, 1933. She thinks wrongly that the tunic is identical with the Greek chiton.

[4] Poulsen, F., *Catalogue of ancient sculpture in the Ny Carlsberg Glyptotek*, no. 462. 2. *Tillaeg til Billedtavler*, pl. 7. Our fig. 2*a*: in the Museo civico at Verona, no. 145. Phot. German. Arch. Institute Rome no. 8302. Our fig. 3: the terracotta in Athens from Phot. German Institute Athens, Nat. Mus., 199. The torso from Samos: Buschor, *op. cit.*, 45, fig. 10. The torso from Prusias: Musées Ottomanes no. 1180. Mendel G., *Catalogue des sculptures Grecques, Romaines et Byzantines* **3**: 339–340, no. 1100, 1914 (with scrinium; head was inserted).

statue, however, is not the direct model for the main Roman type.

While the Hellenistic Greeks and the Romans adopted the general classical models, they failed to keep intact the ideal art of the classical period. Only the outward form remained similar, but the content was altered. They used the classical prototypes for wholly different purposes, expressing new ideas and breathing new meanings into

a *b*

FIG. 4. Statue of youth from Eretria. Athens, National Museum, no. 244. Phot. Alinari, 24295 and Phot. Nat. Mus. Athens.

FIG. 5*a*. Statue of Kleopatra of Delos, *Bull. Corr. Hell.* **31**: 415, fig. 9, 1907.

their artistic creations. This can best be illustrated by the youth of Eretria (fig. 4).[5] Its artist has reshaped the Aischines type in the Hellenistic style. Like its predecessors Sophokles and Aischines, the youth of Eretria has the right arm in a sort of sling made by the upper edge of the mantle. The hand grasps the folds running down from the left shoulder. The left arm—in contrast to the classical type—is hanging at the side and the hand is gathering up some of the lower folds of the mantle. The date of the

[5] Brunn-Bruckmann, *Denkmäler*, pl. 519. Hekler, A., *Bildniskunst*, pl. 51. Arndt-Amelung, *Einzelaufnahmen antiker Skulpturen*, no. 624. Lippold, G., *Kopien und Umbildungen*, 106. Collignon, M., *Statues funéraires*, 283, fig. 175, 1911. Lippold assumes that the youth of Eretria is a variation of an older Hermes type, the "Mercure of Richelieu," with the addition of a mantle taken from daily life. I do not see the connection between the nude and the perfectly veiled statue. The one with the emphasis on the body is classical, the one with the emphasis on the drapery is Hellenistic.

FIG. 5*b*. Tomb relief of Menophila, daughter of Artemidoros of Ephesos, in the type of "Pudicitia." From Ephesos in the Louvre. Phot. Giraudon, No. 2039.

FIG. 5*c*. Tomb relief of Menander. Bibliothèque Nationale, Cabinet des Médailles, Paris. Phot. Giraudon, 1804.

FIG. 6. Votive stele from Panderma, in Museum of Istanbul. Phot. Mus. N. 1604.

statue can be deduced from the similar torso of Dioskurides, a magistrate on the island of Delos. An inscription places his date at 138 B.C. With his torso, the torso of the statue of his wife named Kleopatra were found (fig. 5*a*).[6] She is represented in the type named "Pudicitia," the personification of modesty, on account of her modest and restrained attitude, resembling that of her husband and of the youth of Eretria in the same type. They have the mantles wrapped tightly about their bodies—much more so than the Sophokles and Aischines statues—giving the figure a closed form and modest attitude in contrast to the proud one of the classical statues.

[6] Mayence, F., et G. Leroux, in *Bulletin de Correspondance Hellénique* **31**: 415–419, fig. 9, 1907. Chamonard, in *Exploration archéologique de Délos* **8**: 39–40, fig. 14, and 218–219, fig. 95, Paris, 1922. The statues are dated through the archontate of Timarchos in 138/7 B.C. Buschor, E., *Das Hellenistische Bildnis*, 36 and 38, fig. 9. Best illustrations of Kleopatra: Buschor, *Die Plastik der Griechen*, 103, Berlin, 1936. Bieber, M., *The sculpture of the Hellenistic Age*, 31, fig. 511, New York, Columbia Univ. Press, 1955. Alscher, Ludger, *Griechische Plastik* **4**: 115–117, fig. 49, 1957.

III. REPLICAS ON GREEK TOMBSTONES OF THE HELLENISTIC AND ROMAN PERIODS

The "Eretria" type has been repeated on more than a hundred Hellenistic and Roman tombstones. About eighty or more were made in Attica.[7] About three dozen examples found in Asia Minor are now in the Museum of Istanbul.[8]

[7] Brückner, P., *Ornament und Form attischer Grabstelen*, 93, pl. II, figs. 7–8, 1886. Conze, A., *Attische Grabreliefs* **4**: 23–24, nos. 1820–1822, pl. CCCLXXXVII; 56–59, nos. 1958–1965, pls. CCCCXXI–CCCCXXII; nos. 1970-71, pls. CCCCXXIV (with one woman in the dress of a priestess of Isis, the other as Herculaneum woman); 61, no. 1973, pl. CCCCXXVI; 65, nos. 1989–1991, pl. CCCCXXXI; 66–67, nos. 1994–1995, 1998–2001; 72–75, nos. 2018–2019, 2026–2033, pls. CCCXXXII–III; 77–78, nos. 2040–2044, pls. CCCCXLII, CCCCXLIV–V; 82–83, nos. 2059–2063, 2065–2067, pl. CCCCLII; 85, nos. 2072–2082, 2084, 2085a, 2088–2089, 2098–2103, (2076 wife same type as man, 2080 and 2085a with Herculaneum woman), pls. CCCCLIV–VI, CCCCLVIII–IX, CCCCLXI; 97, no. 2111, pl. CCCCLXII; 98–99, no. 2114, pl. CCCCLXIV (with Herculaneum woman); 99, no. 2117b, pl. CCCCLXV; 100, no. 2118 (wife same type as man); 100–101, no. 2119, pl. CCCCLXVII. Conze speaks of the "typische Haltung," typical attitude of these men on the tombstones. Other examples from the first to the third centuries A.D.: Riemann, Hans, *Keramcikos* **2**, Die Skulpturen, 48–50, no. 46, pl. 15 (palliatus and palliata); 51–53, no. 50, pl. 13,*2*; 54–55, no. 52, pl. 16,1–*2*; 69, no. 76, fig. 3, pl. 20,*3*; 95–96, no. 129, pl. 30 (torso), 1940.

Mühsam, A., *Die attischen Grabreliefs in römischer Zeit*, Dissertation, Berlin, 1936, has treated the Roman-Attic reliefs from the art historical point of view. The excellent work, translated, revised, and well illustrated, appeared in *Berytos* **10**: 51–114, pls. VII–XXIII. For the men in himatia see her pls. VII, 3, XII (with Isis priestess and Herculaneum woman), XVII, 4, XVIII, 3, XXI (with Herculaneum woman), XXIV, 1 and 3 (with Herculaneum woman). Yale University, 1952.

See also the boy standing in the attitude of the Eretria youth between a seated and a standing man, perhaps his father and grandfather, on the relief from the Athenian olive grove, now in Berlin, no. 1462, Conze, **4**: 8–9, no. 1743. Bieber, *Röm. Mitt.* **32**: 118 f and 130, 1917. The Y may point to a neo-Pythagorean school. Picard, C., *Revue archéol.* **23**: 154–155, 1945.

[8] A corpus of the eastern tomb reliefs will be published by Hans Möbius for the German Archaeological Institute in Berlin, based on the material collected by the late E. Pfuhl. Examples in the paper of Pfuhl in *Arch. Jahrbuch des deutschen Instituts* **20**: 55 f., fig. 11; 128, fig. 22; 135, fig. 26, 1905. Stele in Oxford Ashmolean Museum, with "Pudicitia," both in reversed attitude: Horn, R., *Stehende weibliche Gewandstatuen*, 66–67, 78, pl. 25,2. Our fig. 5*c* in the Cabinet des Médailles, Bibliothèque Nationale, Paris. Tomb reliefs with "Pudicitia," see Pfuhl, *op. cit.*, 52, fig. 5; 53, fig. 7; 54, fig. 8; 78, fig. 14; 123, fig. 20; 129, fig. 23; 135, fig. 26.

Many examples are in the Musées Ottomanes. Mendel, G., *op. cit.*, nos. 888, 893–894, 897, 899, 903 (with Hygeia

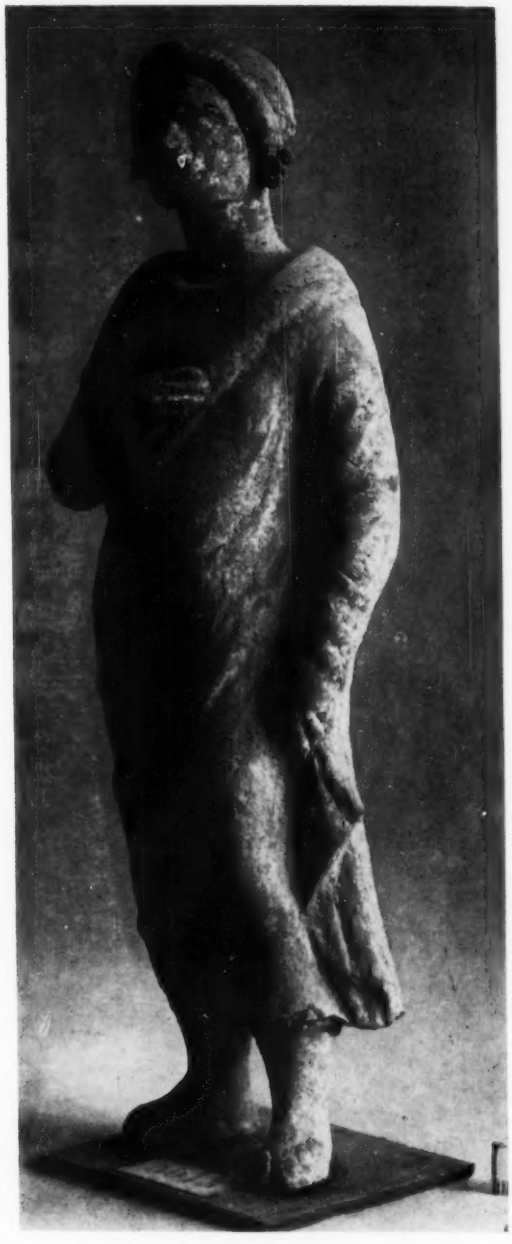


FIG. 7*a*. Youth of comedy, terracotta statuette. Athens National Museum. Phot. German Arch. Inst., Athens, N. M. 1124.

FIG. 7*b*. Father of comedy. Bronze statuette, Metropolitan Museum of Art. Phot. Faraglia.

They seem to have had their beginning in the second century B.C., and to have continued without interruption into the Roman period. Most of the replicas coming from Attica belong to the Roman period, while the eastern tombstones have a larger proportion from the Hellenistic period. The tomb stele of Menander, for example, now in the Bibliothèque Nationale at Paris belongs to the second century B.C. (fig. 5*c*). The stele of Menophila, daughter of Artemidoros of Ephesos, showing her in the type of "Pudicitia" (fig. 5*b*), is contemporary with the one of Menander. The first century stele found in Panderma, now in Istanbul, shows Meidias in the "Eretria" type as a votary at the altar of Zeus (fig. 6).[9]

as Pudicitia), 905 (with Pudicitia), 906, 907, 908 (wife as Herculaneum woman), 910, 911 (wife Pudicitia), 913, 915 (wife Pudicitia), 916, 917 (wife Pudicitia), 918, 919 (wife Pudicitia, daughter smaller Herculaneum woman), 920 (father and three sons in himation, wife larger, daughter smaller Herculaneum woman), 933, 934, 935, 936, 940 (son himation, mother Pudicitia type), 946 (bust), 954, 960 (busts), 991 (funeral banquet, son standing in himation), 1038 (father and son at altar), 1039 (man and wife same type), 1040 (five persons in different sizes same attitude), 1041 (two men in himation; mother in type of large, daughters of small Herculaneum women).

[9] Our fig. 5*b*, tomb relief of Menophila: Bieber, *The sculpture of the Hellenistic Age*, 132, fig. 523. Our fig.

FIG. 8. Ambivius Turpio speaking the prologue of Terence, *Hecyra*. Codex Terentius Vaticanus Latinus 3868, pl. 65.

IV. FABULA PALLIATA

The Hellenistic New Comedy depicted the oldest youth in the same pose and draping as the Eretria youth, and sometimes the father in a similar pose but with the hand higher up and the head veiled (figs. *7a–7b*).[10]

The Roman comedy adopted the characters and situations of Greek New Comedy and with them the Greek outdoor dress, the *himation*. The Latin name for this mantle is *pallium* and, therefore, the Latin comedy is named *Fabula Palliata*. The man in the pallium was the type of the Greek for Plautus. In his *Curculio*, produced in Rome shortly after 194 B.C., the parasite asks everybody to get out of his way:

280 Date viam mihi, noti atque ignoti. . . .
Nec quisquam'st tam opulentus, qui mi obstat in via
Nec strategus, nec tyrannus . . . nec demarchus . . .
290 Tum istic Graeci palliati, capite operto qui ambulant,
Qui incedunt suffarcinati cum libris, cum sportulis,
Constant, conferunt sermones inter sese drapetae
Obstant, obsistunt, incedunt cum suis sententiis, . . .

280 Make way for me, friends and strangers. . . .
Nobody is rich enough to stop me, neither a general nor a tyrant nor a major. . . .
290 And then those Greeks in their mantle (pallia) who walk around
with their heads covered, laden with books and baskets, who stand together, deliver speeches

FIG. 9. Etruscan youth. Bronze statuette, Rome, Museo Nazionale Romano, Antiquario no. 66172. Phot. Emeline Hill Richardson.

5*c*, tomb relief of Menander: Collignon, *Statues funéraires*, 281, fig. 174. Our fig. 6 from phot. Museum of Istanbul, no. 1604; Mendel, *op. cit.* **3**, 41–42, no. 837, Mus. no. 1503.

[10] Pottier et Reinach, *Nécropole de Myrina*, 468 f. pl. XLV, 1. Bieber, M., *Die Denkmäler zum Theaterwesen*, 163, no. 141, pl. 95, 2, Berlin, 1920. *Eadem, The history of the Greek and Roman theater*, 172, fig. 230; 308, fig. 414; 170, note 10, Princeton Univ. Press, 1939. New enlarged and revised edition in the press, 1959.

among themselves, the runaway slaves, stand in our way and walk along with their conclusions. . . .

The tradition of wearing the pallium in comedy was continued. The manuscripts of Terence, in the Vatican and in the Bibliothèque Nationale at Paris, have Lucius Ambivius Turpio, the director of the player's company, dressed in a pallium while speaking the prologues of the *Heautontimorumenos* (the Self Tormentor) and the *Hecyra* (the Mother-in-Law, fig. 8). In my opinion, this constitutes evidence that the tradition of the manuscripts is excellent, and that the original for the manuscripts may be of an earlier period than is now generally assumed.[11]

FIG. 10. Hellenized Etruscan or Roman boy. Princeton University, Princeton, N. J. Phot. Art Museum, Princeton.

FIG. 11. Tomb relief in Galleria Borghese. "Palliati" and "Pudicitia." Phot. Alinari, 27481.

The type representing a man draped tightly in his himation, certainly came to Italy during the Hellenistic period. Plautus, while alluding to the palliati in rather derisive and mocking tone, acknowledges at the same time that they are familiar persons in the streets of Rome. Before the Romans the Etruscans had taken over the Greek types, as Emeline Hill Richardson will

[11] See *Codex Terentius Vaticanus Latinus* 3868, pls. 35 and 65. *Comédies de Terence*, 151, Dessins du Manuscript Latin 7899, Bibliothèque Nationale, Paris, pls. 56 and 107. Jones, L. W., and C. R. Morey, *The miniatures of the manuscripts of Terence prior to the thirteenth century*, figs. 325–326 and 328 (Heautontimorumenos); figs. 584–585 and 587 (Hecyra), 1931. See 201–211 on the different dates given to the originals. The authors on 42–45, 113–116, 203–204 deny the dependence of the miniatures on actual stage representations, because the archetype of the miniatures is late. But in the early Codex, certainly built on still earlier prototypes, masks, garments, and outfits have the right forms used in stage representations. Robert, Carl, *Die Masken der neueren attischen Komödie*, 87–108, Halle, 1911. Bieber, M., *Denkmäler zum Theaterwesen*, 170–174, and *History of the Greek and Roman theater*, 204 and 400–403.

FIG. 12. Tomb relief from Via Statilia, in the Museo Nuovo, Rome. "Palliatus" and "Pudicitia." Phot. German Archaeological Institute, Rome, no. 29.172.

FIG. 13. Baebia, wife of Roman Proconsul Lucius Valerius Flaccus. From Magnesia in the Museum of Istanbul. Phot. Mus. Istanbul, no. 273.

show in her forthcoming study of Etruscan bronzes. I owe thanks to her for the photograph of the little bronze in the Museo Nazionale Romano delle Terme (fig. 9), and to Frances Jones for the bronze statuette of a boy, about twelve years old, in the Art Museum of Princeton University (fig. 10).[12] Emeline Hill believes the latter to be Hellenistic-Roman, while George Hanfmann prefers to classify it as Hellenistic-Etruscan. The left arm is brought farther back-

[12] Bronze statuette, Antiquario, Museo Naz. Romano delle Terme, no. 66172. The Princeton Bronze: Hanfmann, George G. M., An Etruscan bronze, in *Record of the Museum of Historic Art*, Princeton University, **2**: 1, 4–11, figs. 1–4, 1943, H.O.346m. Hanfmann describes it rightly as showing "transition from late Etruscan to early Roman," but decides for late Etruscan, while Emeline Hill Richardson decides for early Roman. Hanfmann calls the dress toga, although he quotes Isidorus, *Origines*, XIX, 24, 17: "Praetexta puerile est *pallium* quo usque ad sedecim annos pueri nobiles sub disciplines cultu utebantur." Two similar terracotta statuettes have been found in Calvi. Alda Levi, *Le Terracotte figurate del Mus. Naz. di Napoli*, 139 f. nos. 628 and 630, pls. V–VI, Florence, 1926.

wards than in the earlier Greek and later Roman examples.

V. REPLICAS ON ROMAN TOMBSTONES IN ITALY

The type of a man draped tightly in a pallium has been represented frequently on Roman republican and early imperial tombstones. Vessberg has collected a good many such examples,[13] but many others also exist. It is interesting to note that on a number of occasions the man in the himation is represented together with the "Pudicitia" type, just like Dioskurides and his wife Kleopatra (fig. 5*a*).[14] Such are the tomb reliefs in Villa Borghese (fig. 11) and in the Museo Nuovo (fig. 12), dated 80–50 B.C.[15] Another example of such a combination in portrait statues of about the same period came from Magnesia. Here was found an over-life-sized torso of a palliatus, with a cavity for the head, supposedly of Lucius Valerius Flaccus, who was an hypatos, proconsul of the province of Asia in 62 B.C. Together with this palliatus were statues of his wife Baebia (fig. 13) and his daughter-in-law, Saufeia (fig. 14), wife of his son Lucius Valerius Leukios, represented on the same scale.[16] Both statues

FIG. 14. Saufeia, daughter-in-law of the Roman Proconsul Lucius Valerius Flaccus. From Magnesia in the Museum of Istanbul. Phot. Mus. Istanbul, no. 156.

FIG. 15. Statue of a woman in the type of Saufeia. From Magnesia in Istanbul. Phot. Mus. Istanbul no. 170.

[13] Vessberg, Olof, *Studien zur Kunstgeschichte der römischen Republik*, pls. XXV–XXXI, XXXV-XLII, dating from about 100–20 B.C. *Cp.* also, Schweitzer, B., *Die Bildniskunst der römischen Republik*, figs. 113–115, 181, 200. Arndt-Bruckmann, *Porträts*, pl. 251 (= Vessberg, pl. XLII, 1): C. Septumius Sabinus, quattuorvir iure dicundo, found in Vulci, now in Copenhagen. Mustilli, *Il Museo Mussolini*, pl. 119, no. 461: Three palliati, one woman Pudicitia, one with hairknot of Octavia. *Encyclopédie phot. de l'Art*, TEL **3**, Louvre, 260, B: Maximus, husband of Domitia, to whom the stele is dedicated.

[14] See note 6.

[15] Galleria del Casino in the Villa Borghese, no. CLXXXVIII: Wilson, L., *The clothing of the Romans*, 152, pl. LXXXVIII, fig. 96. Phot. Alinari, 27481. Museo Nuovo: Mustilli, *Il Museo Mussolini*, 102, no. 9, pl. LVI, fig. 255. Bieber, M., *Entwicklungsgeschichte der griechischen Tracht*, 52, pl. 49, 1. Schweitzer, *op. cit.*, fig. 114. Vessberg, *op. cit.*, 186–188, 211–212, 249–250, pls. XXVII–XXVIII. Horn, R., *Stehende Weibliche Gewandstatuen*, pl. 40, 1. Found on Via Statilia.

[16] Kohte, J., and K. Watzinger, *Magnesia am Maeander*, 209-210, figs. 212–213. Watzinger compares it with Sophokles and Aeschines, dates in the early imperial

FIG. 16. "Pudicitia," statue from Aphrodisias. Museum of Istanbul. Phot. Mus. Istanbul, no. 185.

FIG. 17. "Pudicitia," statue in the Louvre, Paris. Phot. Giraudon, no. 1148.

FIG. 18. "Pudicitia," statue in the Vatican Museum. Phot. Anderson, no. 1440.

display variations of the "Pudicitia" type, with Baebia supporting the left and Saufeia the right elbow with the other hand. These two variations were used for other Roman statues, most of them

period, and names the dress toga. A portrait head was inserted in the socket. H.2.0m. without head. Same type of imperial period below life size, both hands covered (like Ostia, our fig. 50), *ibid.*, fig. 213. Baebia, Saufeia, and other "Pudicitia" types, see *ibid.*, 198–201, figs. 198, 199, 200 (with head preserved, now in Berlin), 202, 209. Mendel, *op. cit.*, **2**: 273–275, no. 550 (Baebia); **3**: 23–27, no. 822 (Saufeia) and no. 823 = our figs. 13–15. H.2.07m. Alscher, *op. cit.*, 129–130, fig. 58. In all the statues the heads and sometimes hands and one foot are worked separately, in order to be able to use the prescribed type for different likenesses, attributes and individual footwear.

FIG. 19. Tomb statue of an old woman. Pompeii, Via delle Tombe. Maiuri, A., Pompei, fig. 113.

in the attitude of Saufeia, for example, on one from Magnesia (fig. 15); on a statue from Aphrodisias in Istanbul (fig. 16); one in the Louvre (fig. 17); and one in the Vatican (fig. 18).[17] It

[17] See for the type of Pudicitia: Hekler, A., *Römische weibliche Gewandstatuen*, in Münchner archäologische Studien Furtwängler gewidmet, 123–124, 159, 167, 177, 231 Type no. XLV. Collignon, M., *Statues funéraires*, 289–291, figs. 181–183. Amelung, W., *Skulpturen des Vatikanischen Museums* **1**: 36 to Braccio nuovo, 23 (our fig. 18). Bieber, M., *Griechische Kleidung*, 67–68 from Aphrodisias, pl. XXXIII, 2–3, (our fig. 16); *Entwicklungsgeschichte der griechischen Tracht*, 52, pl. 49, 1 and 3; *Sculpture of the Hellenistic Age*, 132–133. Horn, R., *Stehende weibliche Gewandstatuen in der Hellenistischen Plastik*, Ergänzungsheft to *Römische Mitteilungen* **2**: 65–66, 78, 80–82, pls. 23, 1–2; 25, 2; 40, 1–2; 41, 1–2, 1931. Lippold, G., *Kopien und Umbildungen griechischer Statuen*, 217–218, 1923. *Musées de l'Algérie* **15**, *Musée Alaoui*, Supplement, pl. 28, no. 1, 1897. Calza, Raissa, in *Notizie degli Scavi*, 234–237, figs. 10–12, 1942. The type is also used for terracottas: Kleiner, *Tanagrafiguren*, 15. Ergänzungsheft des Deutschen arch. Instituts, 162–164, 178, 232, pl. 57 *b–c*, 1942.

FIG. 20. Tomb relief of an old woman, type of "Pudicitia." Naples, Museo Nazionale. Phot. German Arch. Inst., Rome, no. 31,2973.

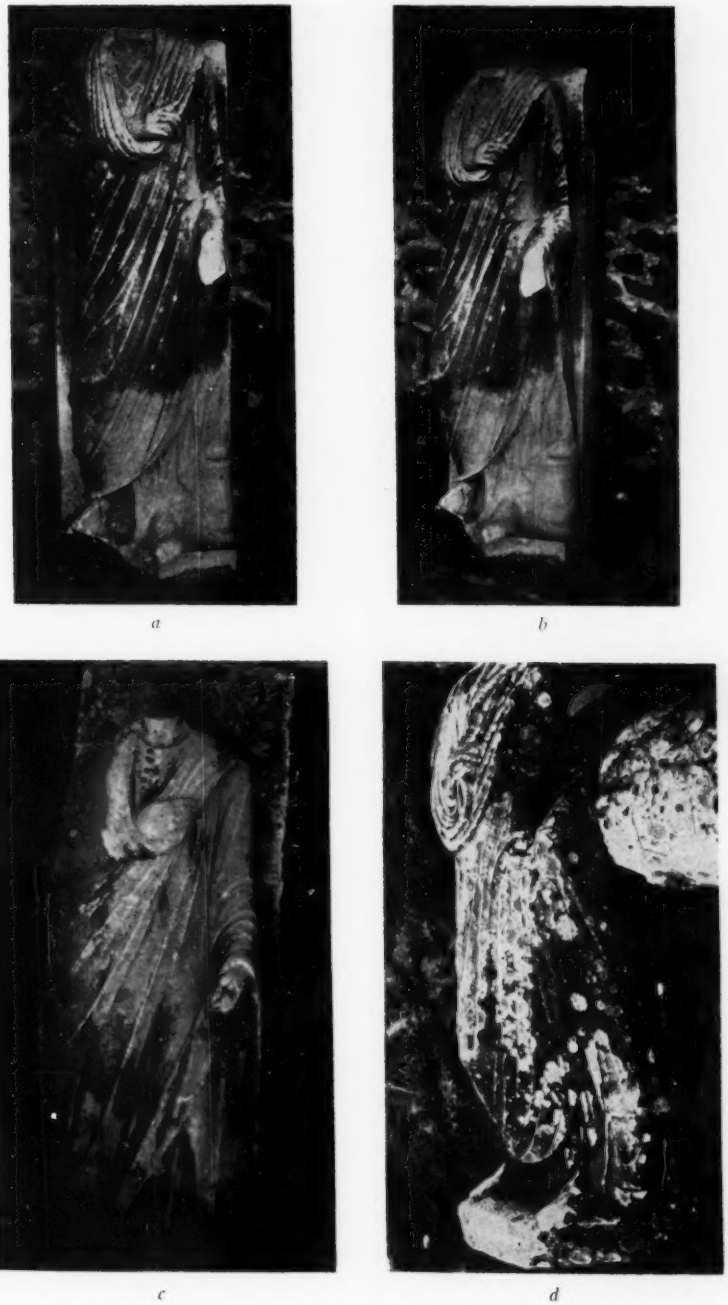
seems that the type was unknown in Athens, while it is used on tomb reliefs in Asia Minor (fig. 5*b*) and in Rome (figs. 11–12). The later Roman republican period applied it particularly to older women in statues, for example, in Pompeii (fig. 19), and both variations on tomb reliefs, for example, in Naples (fig. 20) and in Beneventum (fig. 21).[18]

In contrast, tombstones represented old and young men alike in the draping and the attitude of the palliatus. Sometimes the men actually wear the toga, but it is draped and the arms are held as in the palliati. Many examples are found on the Via Appia (figs. 22*a–d*). However, most of the men on tombstones actually wear the pallium. A number of Roman examples are on display in the outer courtyard of the Museo Nazionale Romano (fig. 23*a–c*),[19] while many are to be

FIG. 21. Tomb relief of an old woman in the type of "Pudicitia." Beneventum. Phot. German Arch. Inst., Rome, no. 32,198.

[18] Our fig. 20 in Naples. Phot. German Inst. Rome, 31, 2973. Our fig. 21 in inversed attitude, in Beneventum Phot. Inst. Rome, 32, 198. For other examples of Pudicitia on tombstones with the palliatus see my list of eastern tomb reliefs in note 8. For other Roman examples, see my note 13, my fig. 28 and Vessberg, pl. XLI, 3, wife of Vibius. Relief in Aquila: Romanelli, P., in *Bulletino comunale* **71**, appendix to vol. 14 (1945), 3–8, figs. 1–3, 1943–1945, one woman in Pudicitia type with five palliati. Romanelli dates it in the early Augustan period. Romanelli dates it in the right period, although he says that he would date it as republican if it had been found in Rome. But the type continues in Rome as well as in the provinces into the imperial period. Tombstone of Lysandra, mother of three children in Venice: Guarducci, M., *Rivista del Arte*, 9, 30–31, no. V, 1942. Portrait statues of Roman ladies in the Pudicitia type: Hekler, *Römische weibliche Gewandstatuen*, in Münchener Archaeologische Studien, 231, no. XLV, 1909. Maiuri, A., *Pompei*, figs. 100 and 113, 1951. Tomb statue of an old lady near the Via delle Tombe = our fig. 19. Near her a torso of the Dioskurides type has been found. For the "Pudicitia" type on Roman sarcophagi, see Lawrence, M., *Amer. Jour. Archaeol.* **32**: 421–424, figs. 3–4, 6–8, 1928.

[19] Our fig. 23: Vessberg, *op. cit.*, 197, pl. XXXVI, 2. Dated *ca.* 30 B.C. Mus no. 105054. Phot. German Institute Rome, 37, 812–814. Three others are at the back wall of the garden court. Our three examples, figs. 22*a–d* from the Via Appia, phots. German Inst. Rome, 34, 1917–1920. Others in Rome, from the tomb of the Scipios: Phot. Inst. Rome, 37, 948. In Villa Doria Pamphili, phot. Inst. Rome, no. 21719. Accademia di San Luca, Garden court in Via della Stamperia. Nearby a "Pudicitia." Vatican, Giardino della Pigna, in niche. Palatine, opposite the entrance from Palace of Tiberius to the cryptoporticus leading to the house of Livia. Villa Medici: Cagiano de Azevado, M., *Le Antichità della Villa Medici*, 45, nos. 17–20, fig. 18. For lists of republican and early imperial tombstones with busts, see note 20 and appendix to notes 19–20, p. 416.

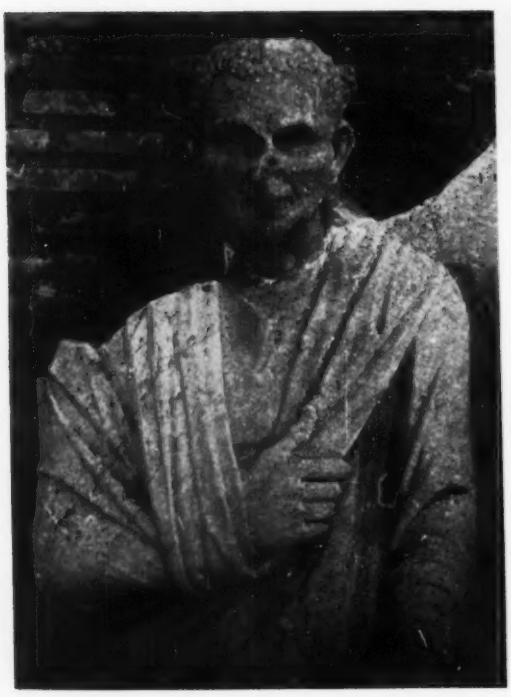


a *b*

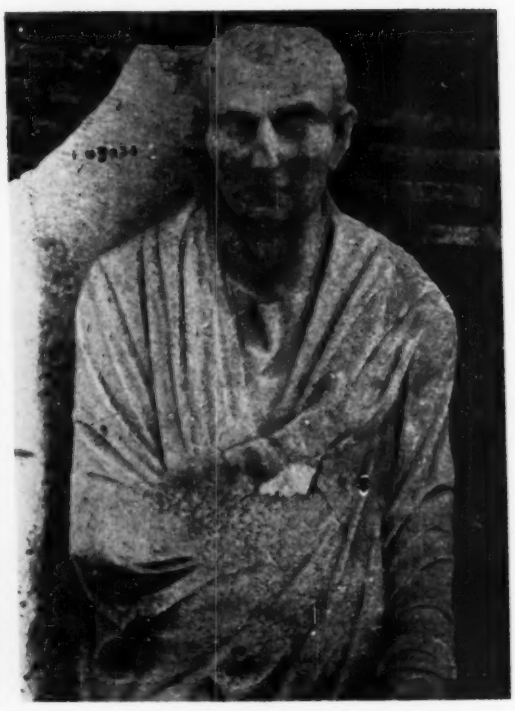
c *d*

FIG. 22. Three tombstones of men in the toga arranged like a pallium. On the Via Appia. Phots. German Arch. Inst. Rome, 34, 1917–1920.

a



b



c

FIG. 23. Tombstone in garden court of Museo Nazionale Romano, no. 105054. Phots. German Arch. Inst. Rome, 37, 812–814.

FIG. 24. Tomb relief in Capua, Museo Campana. Phot. German Arch. Inst. Rome, no. 4685.

FIG. 26. Tomb relief of four persons. Treviso, Villa Reale d'Ossora. Phot. German Arch. Inst., 33,565.

FIG. 25. Tomb relief of five persons. Avellino, Palazzo Municipale. Phot. German Arch. Inst., no. 38, 1246.

FIG. 27. Tomb relief of Firmius and his wife. Treviso, Villa Reale d'Ossora. Phot. German Arch. Inst., no. 33,560.

FIG. 28. Tomb relief of two men and "Pudicitia." Ascoli Piceno. Phot. German Arch. Inst., no. 37,412.

found in the provinces of Italy outside Rome. Some of these may again be men in the toga, arranged in the imitation of the pallium. Such are the full figures on the tomb relief in the Museo Campana in Capua (fig. 24); the busts on those in Avellino (fig. 25), Treviso (figs. 26–27), Ascoli Piceno (fig. 28), and the collection of tomb reliefs built into the campanile of the Cathedral at Beneventum (fig. 29).[20] Females often appear with the men, but the "Pudicitia" shows up on rare occasions only, as in Ascoli Piceno (fig. 28). In fact, the "Pudicitia" type seems to have disappeared during the time of the Emperor Augustus. It gives way gradually to other Greek types, especially the two women from Herculaneum and to free variations of other Greek types.[21]

The tomb relief of the Furii family in the Lateran Museum, dated around 30 B.C., affords an example of this trend during the early Augustan period (fig. 30).[22] The two men are presented in the familiar pose of the palliatus. But the three women, with hairknots over their foreheads—imitative of the hair styles worn by Octavia and Livia—have freer and diversified movements than the men. Similarly, the tomb relief

[20] Our fig. 24 in Capua, Museo Campana. Phot. German Inst. Rome, 4685. Fig. 25, Avellino, Palazzo Municipale. Phot. German Inst. Rome, 1938, 1246. Four persons in same attitude. In center boy with cock. Probably three generations of same family. Fig. 26, Treviso, Villa Reale d'Ossora. Phot. German Inst. Rome, 1933, 565. Three men and woman in same attitude. Woman has parted hair. Fig. 27, *ibid.*, phot. German Inst. Rome, 33, 560. Firmius and wife. Same attitude. Wife parted hair. Fig. 28, Ascoli Piceno. Phot. Inst. Rome, 1937, 412. Two palliati. Woman, probably wife and mother Pudicitia type. Fig. 29, Campanile of the Cathedral in Beneventum. Phot. Inst. Rome, 1930, 601. Sixteen half-figures made into a modern frieze, probably from seven to eight tombstones. For other examples, see appendix to notes 19-20, p. 416.

[21] See examples on Attic tombstones in note 7, and on republican tombstones in our figures 25–27, 30.

[22] Tomb relief of the Furii: Lateran Museum no. 723. Benndorf, O., and R. Schöne, *Die antiken Bildwerke des Lateranensischen Museums*, no. 467. Vessberg, *op. cit.*, 197–198, 232, pl. XXXIX, 2. Goethert, F. W., *Die Kunst der römischen Republik*, 38–40, Berlin, 1931. West, R., *Römische Porträtplastik*, 1: 141–142, pl. XXXV, fig. 149. Phot. Anderson 24229.

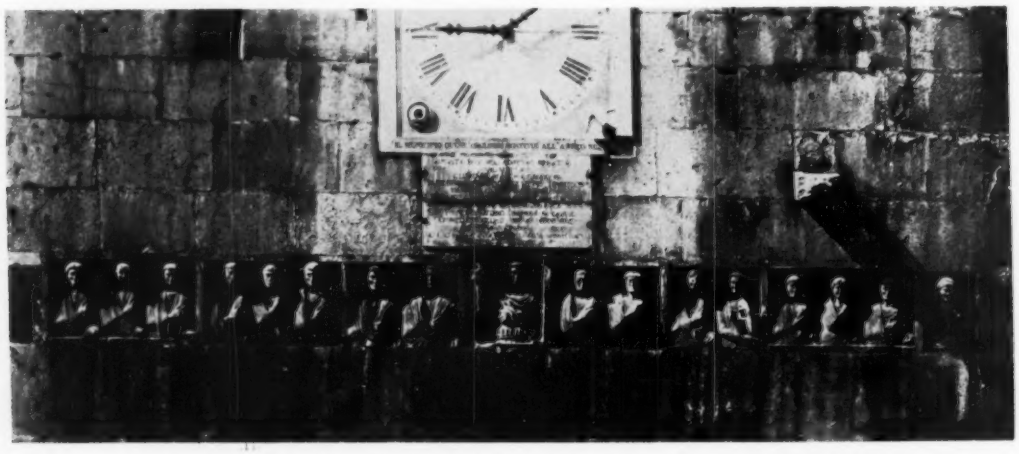


FIG. 29. Frieze made of several tombstones on the Campanile of the Cathedral of Beneventum, below the clock. Phot. German Arch. Inst., no. 30,601.

FIG. 30. Tomb relief of the family of the Furii, Rome, Lateran Museum, no. 467 (866). Phot. Anderson, 24229.

found outside of Porta Maggiore (fig. 31),[23] probably belonging to the baker Eurysaces, pictures the wife drawing up her dress with the right hand inside the mantle, a motif reminiscent of "Livia" or rather Julia on the Ara Pacis. However, the man maintains the severe unchanged attitude of the palliatus, though the mantle may be the toga. Again, a tomb relief in Trieste (fig. 32)[24] has the wife turning to embrace her husband, who, clothed in a pallium, stands stiffly

FIG. 31. Tombstone of the baker Eurysaces and his wife. Rome, outside Porta Maggiore. Phot. German Arch. Inst., Rome, no. 33,749.

FIG. 32. Tombstone of husband and wife. Trieste. Phot. German Arch. Inst. no. 37,660.

[23] This relief, found outside Porta Maggiore, might belong to the neighboring tomb of Eurysaces. Early Augustan. Fig. 31 from phot. Inst. Rome, no. 1933, 749. Vessberg, *op. cit.*, 195, pl. XXXIV, 3.

[24] Fig. 32 from phot. Inst. Rome, 37, 660. Trieste, Museo civico di storia ed arte. Left hand originally held a scroll. Conze, *op. cit.* 4: 94, no. 2103, pl. CCCCLIX, considers the relief to be Attic.

FIG. 33. Tombstone in Pozzuoli, Amphitheater. Phot. German Arch. Inst., Rome, no. 37,643.

in a frontal position. A relief from Pozzuoli, the date of which can be deduced from the "Agrippina" hairdress of the Claudian period, seemingly has the wife trying to imitate the position of the right hand of the husband (fig. 33).[25]

[25] Fig. 33 from phot. Inst. Rome, 37, 643, found in the Amphitheater of Pozzuoli.

VI. ROMAN STATUES

Beside the many tomb reliefs, there are also many statues in the round of the Romanus palliatus. There are nine examples from Rome collected in the Museo Nazionale Romano delle Terme Diocleziano.[26] Two torsos are in the

[26] Museo Nazionale Romano, Aula VIII, a headless torso. Aula IX, Mus. no. 121531. Head and hands were inserted. Two in the upper niches of the façade of the museum looking to the outer garden court. Four in the Exedra, in the niches right and left to the entrance from

a

b

FIG. 34*a–b*. Statue of the Claudian period. Avignon, Musée Calvet. Phot. German Arch. Inst., Rome, no. 33, 403–404.

Giardino della Pigna of the Vatican.[27] Others can be seen in the Museo Nuovo,[28] on the Largo Argentina near the republican temples[29] not far from the Piazza with the monument for the brothers Carioli, and on the Viale leading from the Spanish staircase to the Pincio. The Villa Celimontana possesses two such statues, one dated about 100 B.C. and the other 80–50 B.C.[30] Another is in the courtyard of the Palazzo Rospigliosi, dated about 50 B.C.,[31] while another is in a private collection at Rome.[32] The Museo Civico of Chiusi has four statues, belonging to the second half of the first century B.C.[33] Four statues are in the Glyptothek Ny Carlsberg in Copenhagen, two of which are products of the early part of the first century, while the remaining two are of the period from 70 to 40 B.C.[34] A statue in the Louvre dates from the time of Caesar.[35] To the Claudian period belong a statue in Avignon (figs. 34*a–b*)[36] and another in the Villa Doria Pamphili (fig. 35).[37]

Torsos with empty sockets for the heads to be inserted are located in the Garden of the Chiostro di Michelangelo of the Museo Nazionale Romano. The upper court of the Museo Nazionale di Napoli contains two statues, also without heads. They have boxes for books next to the left legs. From

FIG. 35. Statue in the Villa Doria Pamphili, Rome. Phot. Arch. Inst., Rome, no. 7417.

Alexandria probably comes the statue of Julianus, now in the Egyptian museum of the Vatican (fig. 36).[38] This statue, carved in the Claudian

the garden court. Three torsos and one statue with head. The latter no. 952. Wilson, L., *The Roman toga*, 39–42, figs. 9–10; *cp.* fig. 11 in a private collection; and *The clothing of the Romans*, 39–40, fig. 31, pl. XXIV. Torso with cavity for head in the garden of the Chiostro of Michelangelo, No. 960: Mai, Felletti, B. M., *Museo Nazionale Romano, I Ritratti*, 32, no. 42, fig. 42. Republican before middle of first century, with scroll in left hand and a round bookcase near his left leg. Mrs. Felletti calls the pallium "toga exigua."

[27] Giardino della Pigna, two torsos in niches. Amelung, W., *Skulpt. vat. Mus.* 1: 845–846, no. 91, pl. 99.

[28] Mustilli, *Il Museo Mussolini*, 8, no. 15, pl. VI, fig. 20.

[29] Standing upside down between the northern and the republican temples. Torso only preserved to upper legs.

[30] Vessberg, *op. cit.*, 175–180, pl. XXIII, 1–2 and 186, pl. XXIX, 2–3.

[31] Vessberg, *op. cit.*, 190, pl. XXXII, 4.

[32] Wilson, *The Roman toga*, fig. 11. Phot. Moscioni.

[33] Chiusi, Museo Civico. Vessberg, *op. cit.*, 240–241, nos. 6–9, pl. LXXXV, 1–4.

[34] Poulsen, F., *Cat. of sculptures in the Glyptothek Ny Carlsberg*, nos. 496 and 528. *Billedtavler*, pls. XXXVII and LVIII. Hekler, A., *Bildniskunst*, pl. 129*a*. Arndt-Amelung, *Einzelaufnahem antiker Skulpturen*, nos. 2019 and 4771–2. West, R., *op. cit.* 1: 41–44, pls. VIII–IX, figs. 27 and 29. Vessberg, *op. cit.*, 193 and 220, pl. LVIII. Schweitzer, *op. cit.*, 80–85, E8, figs. 109 and 113.

[35] Louvre, *Catalogue sommaire*, 53, no. 920; West, *op. cit.* 1: 94–95, pl. XXII, fig. 89.

[36] Musée Calvet, Avignon, fig. 34 from phot. Inst. Rome, 1933, 403–404.

[37] Fig. 35 from phot. Inst. Rome, 7417.

[38] I owe phot. for fig. 36, Vatican, to the kindness of Hermine Speier. Botti-Romagnoli, *Le Sculture del Museo Gregoriano Egizio*, 120, no. 192, pl. LXXXI. Mus. no. 42. The author calls him "philosopher or orator." Probably also to the first century A.D. belong an

FIG. 36. Statue of Julianus. From Alexandria in the Egyptian Museum of the Vatican. Phot. Archivio Fotografico Gall. e Musei Vaticani, no. XXXII-2-31.

FIG. 37. So-called "Mario," statue in the Capitoline Museum. Phot. Anderson, no. 1735.

period, holds a scroll in the left hand. It bears the inscription EIOYLIANOS, the Greek form

unfinished statue in the Entrance Hall of the Museo archaeologico in Florence, with originally inserted scroll in left hand, and two torsos in the upper courtyard of the Museo Nazionale di Napoli. All three lack the heads which were inserted in sockets. Scrinia = boxes for rolled manuscripts beside the Naples figures. A statuette of the early imperial period in Berlin, the head crowned with laurel, book scrolls bound together at his feet (*Beshreibung der antiken Skulpturen*, 156, no. 387, 1891).

a *b*

FIG. 38. The Emperor Hadrian. Statue from Cyrene in the British Museum, no. 1381. Phots. Courtesy of the Trustees of the British Museum, London.

FIG. 39. Statue of a priest. Found in Kyzikos, in the Museum of Istanbul, no. 582.

for Julianus. Another statue, carved in basalt, was found in Erment, south of Luxor, and is now in Cairo. It has been wrongly labeled "a person in a toga." [39]

All these statues have wide mantles with many folds. There can be no doubt that this mantle is the himation, called pallium by Plautus, and not the toga. For example, the toga worn by Aules Metelis, the "Arringatore," has the rounded republican Roman form of the toga exigua, an elongated half ellipse with only two corners, while the examples enumerated above have the square Greek shape, with four corners.[40]

FIG. 40. Statue of Neokles, son of Herodoros, ruler of Georgippias, on the Black Sea. Found in Anapis. From Blavatsky, *The Art of the Northern Coast of the Black Sea*, fig. 61.

[39] E. Driorton, *Encyclopédie photographique de l'Art. Le Musée de Caire*, ed. TEL, 15 f., fig. 219: "Personage en toge," 1949. Graindor, P., *Bustes et statues portraits d'Egypte Romaine*, Univ. Égypt. Cairo, 102–104, no. 47, pl. XLI*a*, 1930. Palliatus wearing the calceus patricius, holding a scroll in his left hand; a bookcase and above it a bunch of scrolls beside his right leg. See also, *ibid.*, the later statues 93–94, no. 42 (6022), pl. XXXV (*cf.* note 46) and 101-102, no. 46, pl. XL (limestone). Graindor has recognized the himation instead of the toga.

[40] Arringatore: Hekler, *Bildniskunst*, pl. 129*b*. Arndt, *Porträts*, pls. 86–88. Vessberg, *op. cit.*, 171–172, pl. XIX. Giglioli, *Arte Etrusca*, pl. 369. For the toga of

The Emperor Augustus prescribed the use of the toga for public appearances of the imperial family and officers of high positions. But the decree was not always followed. Thus, in the early second century we see Trajan wearing the pallium, represented by a statue found above Tivoli. It is now in Wannsee, near Berlin, in the possession of Hugo Vogel.[41] The Greek letters between the feet indicate the work of a Greek artist, so many of whom were in the service of the Romans.[42] To the same period probably belongs the statue of a man with a modern head, the so-called Mario in the Capitoline Museum (fig. 37).[43] A box holding scrolls stands behind his left foot. The philhellenic Emperor Hadrian, or perhaps one of his priests, is shown wearing a pallium in a statue from Cyrene, now in the British Museum (fig. 38*a–b*).[44] The left hand holding a sprig does not belong; the hand probably held a scroll. The treatment of the drapery, particularly as seen from the side, is typically Hadrianic. The costume looks pleasant, but there is no sharp, clear-cut division between the folds running from both the inside and the outside of the arm. Similar in style and probably of the same period, is the statue of an imperial priest, wearing a diadem with small busts, from Kyzikos in Constantinople (Istanbul) (fig. 39).[45] The statue in Cairo, portraying a lean man with curly hair may belong to the late Hadrianic or early Antonine period.[46] The statue found in Anapis, a small town on the north coast of the Black Sea, depicts a bearded man with a torque around his neck (fig. 40).[47] The inscription testifies that he is Neokles, the son of Herodoros, and was ruling Georgippias as an archon or prefect in A.D. 187.

To the second century, probably the Antonine period, belong six torsos of palliati, of whom five have empty sockets for inserting portrait heads (figs. 41–46).[48] They have been found on the Athenian Agora, and I owe the photographs to the kindness of Homer Thompson and Lucy Talcott. Interestingly, the folds on the fronts agree with each other, but those on the sides and backs are arranged differently (*cp.* figs. 41*b*, 42*b*, 43*b*, 44*a–b*). This shows that, though the type was continued, only the fronts were copied exactly. The artists worked out the sides and backs more or less arbitrarily. This evidence suggests that Greek copyists did not use all-around mechanical casts, as has been assumed by Gisela Richter.[49] Two of these statues (figs. 41, 43) and the very fragmentary torso lacking his upper part (fig. 45) have a bookcase beside the left foot.

Two statues in the Louvre are thought to be

the Arringatore and models for it, *cp.* Wilson, Lillian, *The Roman toga*, 17–39, figs. 1, 3–7; *The clothing of the ancient Romans*, 36–39, figs. 27–30, pls. XX–XXIII. For the square himation, *cp.* Bieber, *Griechische Kleidung*, 21–24, 80–84, pls. LI–LVIII and our figs. 2*b* and 51.

[41] Blümel, Carl, in Arndt-Amelung, *Einzelaufnahmen antiker Skulpturen*, serie XIV A, no. 3901. Gross, W., *Das römische Herrscherbild* **2**, 2. *Bildnisse Trajans*, 131, no. 59. West, R., *Römische Porträtplastik* **2**: 67, no. 5, pl. XVIII, fig. 67, München, Bruckmann, 1941. Head and statue were found above Tivoli, not together but near each other, and the head fits into the socket.

[42] See Toynbee, J. M. C., Some notes on artists in the Roman world, in *Collection Latomus* **6**: 5 f., 1951. Richter, G., Who made the Roman portrait statues, the Greeks or the Romans? in *Proc. Amer. Philos. Soc.* **95**(2): 184–191, 1951, and *Three critical periods in Greek sculpture*, 37–65, 1951.

[43] Jones, Stuart, *Sculptures of the Capitoline Museum*, 284, Salone no. 14, pl. 69. Phot. Anderson 1735. I owe the phot. for fig. 37 to the kindness of Raissa Calza. The head seems to have been set on the statue in the seventeenth century?

[44] Smith, A. W., *Catalogue of sculpture in the British Museum* **2**: no. 1381, 1900. Bieber, M., *Entwicklungsgeschichte der griechischen Tracht*, 52, pl. 51. Graindor, *Athènes sous Hadrien*, 175. Hinks, *Greek and Roman portrait statues*, 29, pl. 35. *Cp.* the statue, *ibid.* 28, pl. 33, possibly Nerva. Wegner, Max, *Das römische Herrscherbild*, **2**, 3, Hadrian, 100–101, pl. 16a.

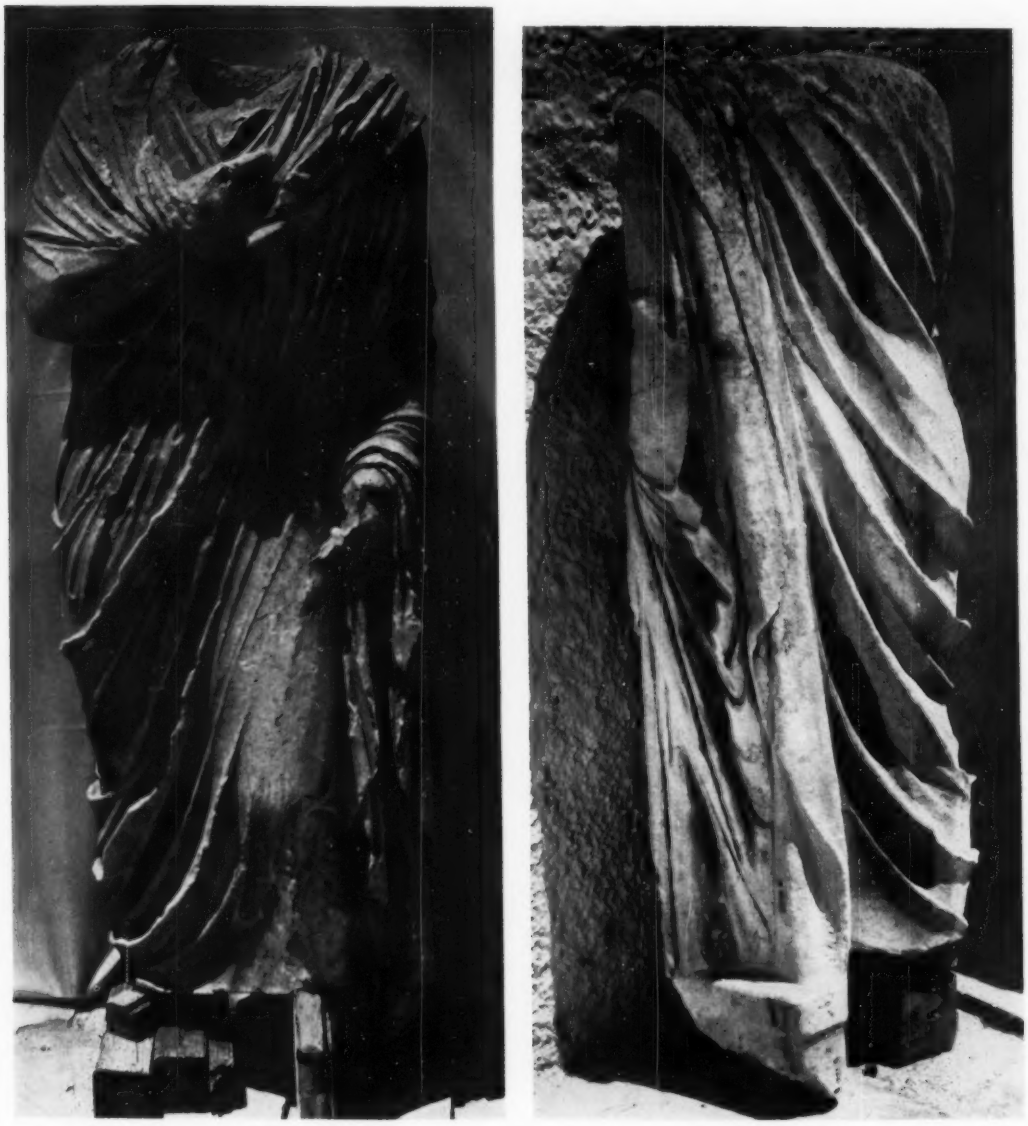
[45] Mendel, *op. cit.*, **2**: 311–313, no. 582.

[46] Kairo Mus. no. 6022. Rubensohn, O., *Arch. Anzeiger* **33–39**: 327, 331, no. 3. fig. 1, 1923/4. Rodenwaldt, G., *Antike Plastik Walter Amelung zum 60. Geburstag*, 209–210, fig. 1, 1928. Graindor, *Bustes et statues portraits d'Egypte Romain*, 93–94, no. 42, pl. XXXV. Phot. Inst. Rome, pl. 116, 2, 1934.

[47] Fig. 40 from phot. German Inst. Rome, 1952, 4, taken for me after Blavatsky, *The art of the northern coast of the Black Sea*, 97, fig. 61, 1947. The Russian text was kindly translated for me by Raissa Calza. Gaidukevich, *The Bosporan Kingdom*, Institute of the History of Material Culture, Academy of Sciences of the U.S.S.R., Moscow-Leningrad, 218–220, fig. 34, 1949. Kubylina, *The sculpture of the Bosporus*, Material and Studies in the Archeology of the U.S.S.R., 171–172, fig. 1, 1–3, and fig. 4, 1, Moscow, 1951. These two references were translated for me by Irwin Scollar.

[48] Found on the Agora by Leslie Shear and Homer Thompson. Shear, *Hesperia* **7**: 323, fig. 8 (our fig. 41), 1938. Inv. nos. 849, 850, 936, 1346, 1347, 1354. Harrison, Evelyn, *The Athenian Agora* **1**: *Portrait sculpture*, 74–78, nos. 57–62, pls. 38–40, Princeton, 1953. A similar torso also found in Olympia, Treu, *Olympia* **3**: *Die Bildwerke*, 251–252, fig. 287, pl. LXII,3. A bookcase is near the left leg.

[49] Richter, *Three critical periods in Greek art*, 441: "Exact copying with the use of casts and the pointing process."



a *b*

FIG. 41. Torso from the Agora Excavations, Athens. Inv. no. S.936. Phots. Lucy Talcott. Front and back.

those of Julian the Apostate, the renegade, who was proclaimed emperor in Lutetia (Paris) in A.D. 360 (fig. 47*a–b*).[50] Some say that this one (fig. 47) was not found in Paris, but possibly brought there via Marseilles; that both probably

[50] The replica (phot. Giraudon, 1876), formerly Musée de Cluny, now in the storeroom of the Louvre. Coarse work. Fig. 47*a–b* from phots. Giraudon, 1877–1878. Louvre, no. 1121. *Catalogue sommaire*, no. 2323. *Encyclopédie phot.* TEL 3: Louvre, 310–311. Bernoulli, *Röm. Ikonographie* 2, 3, 241–247, pl. LIII*a* and *b*. Wulff, O., *Altchristliche und byzantinische Kunst* 1: 152–153, fig. 45.

represent priests and belong to the beginning of the third century. The one formerly in the Musée de Cluny, however, certainly was found in Paris. I believe, moreover, that the style is of a later period, and that the statues are important examples of the classical revival, which, under the leadership of the old Roman nobility, symbolizes a return to pagan religion and Graeco-Roman

a *b*

FIG. 42. Torso from the Agora Excavations, Athens. Inv. no. S.850. Phots. Lucy Talcott. Front and back.

a *b*

FIG. 43. Torso from the Agora Excavations, Athens. Inv. no. S.1354. Phots. Lucy Talcott. Front and back.

art forms.[51] The execution is less baroque than in the statues of the Antonine period. It is dry, but exact. The arrangement at the left side is much better and clearer than those of the Hadrianic statues (figs. 38–39). The ends of the himation are well separated from each other. The two lower corners, marked by tassels, appear one above the other. The one originally in the rear

[51] See Dohrn, Tobias, *Mitteilungen des Deutschen Instituts* **2**: 126–127, 1949. He calls the period Theodosianic Renaissance. The earlier date was suggested to me by Professor Ingholt of Yale University.

hangs higher than that originally in the front, because the folds running from the back are laid over the left wrist. One of the upper corners hangs down from the left shoulder in the back. The other upper corner, coming down in front from the same shoulder, is invisible, because the end is covered by the bulk of the mantle. There are only a few full statues of palliati coming from

a *b*

FIG. 44. Torso from the Agora Excavations, Athens. Inv. no. S.849. Phots. Lucy Talcott. Back and front, three-quarters from left side.

FIG. 45. Right leg and two fragments with book case. From the Agora Excavations, Athens. Inv. no. S.1347.

the late Roman period. The best is the well-preserved statue in the National Museum of Bucharest, dated in about 370–380 by L'Orange.[52] The use of the pallium in public life may have been restricted at this time in favor of the toga.

Four statues of palliati have been found in Ostia (figs. 48–50).[53] Two of them, one in a niche at the Forum (figs. 48*a–b*) and another life-sized on the square of the Corporations (fig. 49) represent the regular type with simple folds. The other two are variations, with both hands veiled, feet nude and a small altar instead of a bookcase (fig. 50).[54] The altar and the covered hands suggest that the figures represent priests. The rich folds seem to give the impression of the actual mantles worn in life (see my model, fig. 51). The general arrangement, however, reminds one of the conventional type.

VII. SARCOPHAGI

The "Sidamara" sarcophagi have many examples of the palliatus, which Morey had called the "Sophokles" type. They begin in the second century and continue into the early fourth. The sarcophagus of Claudia Antonia Sabina (fig. 52*a*)[55] and a sarcophagus in Ostia (fig. 52*b*)[56]

FIG. 46. Torso from the Agora Excavations, Athens. Inv. no. 1346. Phot. Lucy Talcott.

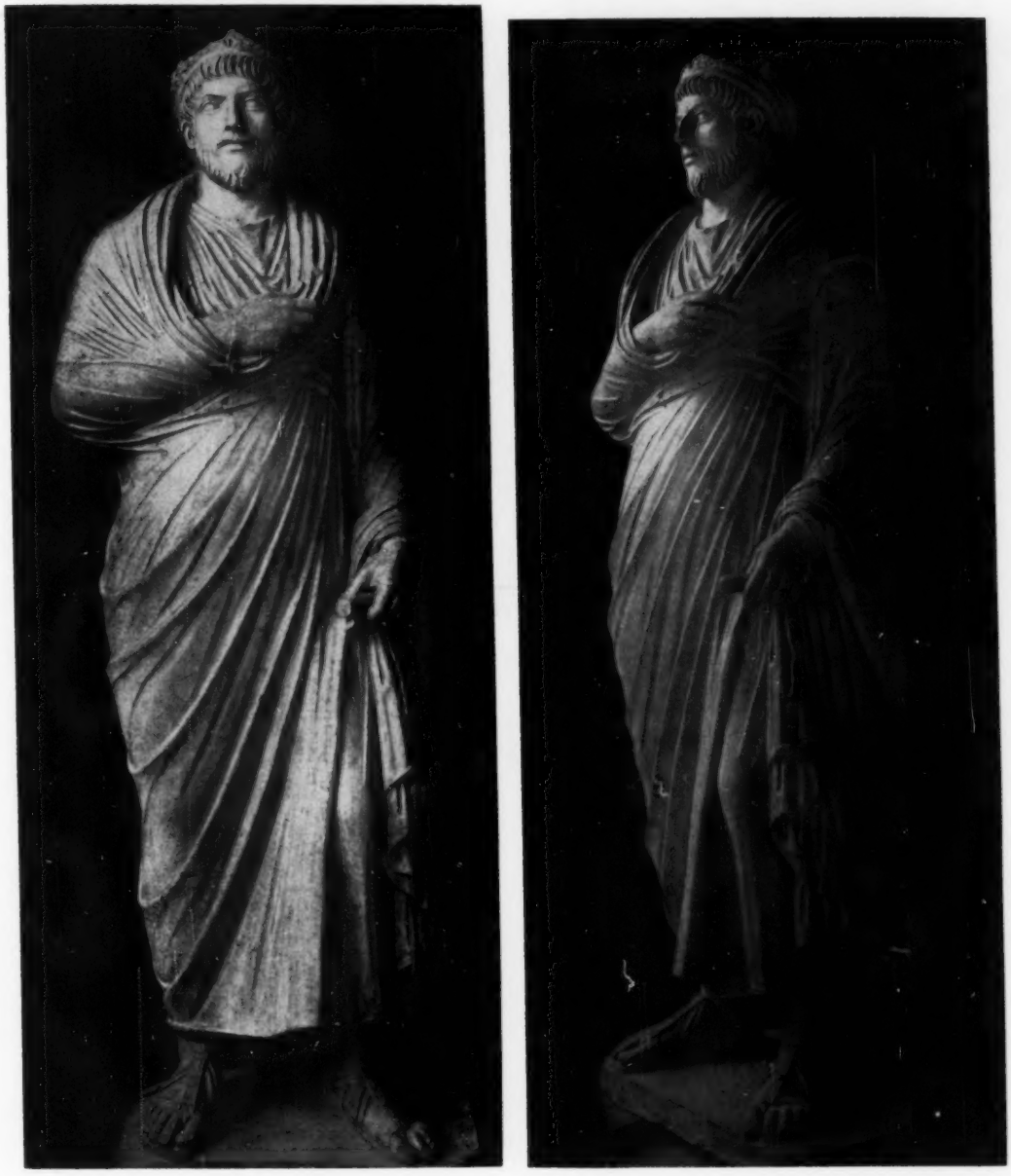
[52] L'Orange, *Studien zur Geschichte des spätantiken Porträts*, 22–26 and 112, Cat. no. 17, fig. 47, 1933. (L'Orange calls the mantle a toga.) A bundle of scrolls stands beside him.

[53] I owe the photos for figs. 48–50 to Raissa Calza. Fig. 48, Mus. No. 1233, formerly in the Castello, Scala 46, now in the niche on the Forum. Fig. 49, Mus. no. 1208, now on Piazzale delle Corporazioni.

[54] Fig. 50 from photograph of Raissa Calza. Found in the Collegio degli Augustali in 1941, Mus. no. 1147, together with a second example, Mus. no. 1111. Calza, Raissa, in *Notizie degli Scavi*, 218–222, fig. 2, 1942. See for a statue in a similar type, found in Rome: Morpurgo, L., in *Notizie degli Scavi*, 202-205, figs. 2–3, 1937. Hekler, A., *Die Antiken von Budapest*, 62, no. 50, fig. 102.

[55] Morey, *Sarcophagus of Claudia Sabina, Sardis* **5**: fig. 11.

[56] Lawrence, Marion, in *Memoirs of the American Academy in Rome* **20**: 141–142, fig. 29 (= our fig. 52*b*), 1951, found east of the Caserma dei Vigili. Phot. given to me by Raissa Calza.



a *b*

FIG. 47. So-called Julian the Apostate, Louvre, Paris: Louvre, no. 1121. Phot. Giraudon, 1877–1878.

a

b

FIG. 48. Statue in Ostia, in a niche of the Forum, formerly Castello, Scala, no. 46. Mus., no. 1233. Phot. Vasari. Courtesy of Direzione Scavi Ostia Antica.

FIG. 49. Statue in Ostia, on the square of the Corporations. Mus. no. 1208. Courtesy of Direzione Scavi Ostia Antica.

are two of more than a dozen examples (see my list).[57] But the same type was also used on many other Roman sarcophagi in full figures of men as well as of boys.[58]

[57] Examples for palliati on Sidamara sarcophagi: Morey, *op. cit.*, his types 1 and 17, figs. 105 and 120; 60, fig. 134. *Cp.* also, frontispiece, figs. 9 and 11. Fragment showing Christ in Berlin, fig. 25 (our fig. 58). Morey calls the type "philosopher," 12, and "type of Sophokles," 68. Lawrence, Marion, in *Amer. Jour. Arch.* **32**: 431, fig. 12 (Salona) and 432, fig. 13 (from Porta Salaria), 1928, *Eadem* in *Memoirs Amer. Acad. in Rome* **20**: 127–128, fig. 9; 133–135, fig. 19 (Beirut); 141–142, fig. 29 (Ostia); 143–144, fig. 31 (Rome); 151–152, fig. 40 (Ostia); 162, no. 1 (Smyrna), 1951. She also names the type "standing philosopher" and names the original as "Sophokles of the Lateran." Mustilli, *Musco Mussolini*, 150–151, no. 6, pl. XCVI, no. 356–7 (from Via Ostiense). Antonine period.

[58] Unfinished sarcophagus in Copenhagen, Glypt. Ny-Carlsberg. F. Poulsen, *Catalogue of ancient sculpture*, no. 780*a*. Ill. in 2nd. vol. of *Tillaeg till Billedtavler*, pl. III. Lawrence, M., in *Amer. Jour. Archaeol.* **32**: 432, figs. 12–13, 1928; *Amer. Jour. Archaeol.* **62**: 285 and 287, pl. 78, figs. 15, 16, 20, 1958. Rodenwaldt, G., in *Arch. Anzeiger* **45**: 168–178, figs. 1, 3–4, 1930. Bellerophon

FIG. 50. Statue with hands veiled, a priest? Collegio degli Augustali, Ostia. Mus., no. 1147. Phot. Raissa Calza.

Busts of the same type were often fitted into round medallions. An example is in the Garden of Palazzo Colonna (fig. 53).[59] Rounded busts of this type, with sockets for fitting portraits of

FIG. 51. Model of a palliatus with veiled hands. Phot. Bruno Schroeder.

the deceased, have been found in great numbers on the island of Thera.[60] An early example from

sarcophagus Athens, Proitos. Robert, C., *Sarkophagreliefs* **2**, 3: 146–147, pl. L, no. 138. Weitzman, K., *Greek mythology in Byzantine Art*, 178–179, fig. 228. Gilbert Picard, *Les Religions de l'Afrique antique*, 214 f., pl. IX, 1954. Picard rightly explains the child on this sarcophagus as a boy. He is standing before a curtain with a scroll in his hand and a bundle of scrolls at his feet. It cannot be a girl as others believed. I do not know of representations of girls with scrolls and scrinium. The dress is too short for a girl.

[59] Fig. 53 from phot. German Archeological Institute, Rome, no. 33, 1717. A similar medallion in the storeroom of the Vatican: Kaschnitz-Weinberg, G. von, *Sculture in Magazzini del Vat.*, 293, no. 717, pl. CXI. Time of Alexander Severus.

[60] Hiller von Gaertringen, Thera, **1**: *Die Insel Thera*, 227, fig. on p. 228r, Berlin, 1904. One of the six busts found in the Basilike Stoa. H.0.46, Br. 0,52m. *Cp.* **3**: 28, fig. 16. Half figures with sockets for heads, *ibid.* **3**: *Stadtgeschichte von Thera*, pl. 23 and 192–193, figs. 207–209. Fig. 207 has three palliati and one Pudicitia. Dated Antonine period and later. *Cp.* Ross, *Inselreisen* **3**: 30, and *Arch. Aufsätze* **1**: for busts with inserted heads from a cemetery near Karterades on Thera. Half figures from Thera in Athens, Kabbadios Kat. no.

FIG. 52*a*. Sarcophagus of Claudia Antonia Sabina. Morey, *Sardis* **5**: fig. 11.

FIG. 52*b*. Palliatus from a "Sidamara" sarcophagus. Ostia. Found near the Caserma dei Vigili. Mus. no. 1283. Phot. Vasari. Courtesy Direzione Scavi Ostia Antica.

about 50 B.C. is in Oslo. A late one from the fourth century A.D., found in Chichester, Sussex, is in Princeton.[61] On the sarcophagi of the third century the medallions with such busts are often held by flying cupids, like the one of Annios Katos in the Lateran Museum (fig. 54).[62]

This type had also been accepted for women. Thus a sarcophagus in the Capitoline Museum is almost identical with that of Annios Kato (fig. 55).[63] It has the same medallion carried by cupids, the same masks below the medallion, the same Okeanos and Tellus in reclining positions, and the same mourning genii of sleep and death at the corners. The portrait, however, is

FIG. 53. Bust fitted in a round Medallion. Palazzo Colonna, Rome. Phot. German Arch. Inst. Rome, no. 33,1717.

780. Benndorf, O., in *Oesterr. Jahreshefte* **1**: 1–5, fig. 2, 1898. See also the bust of Anima Hilara in the same attitude, on the cover of the Endymion sarcophagus in the Metropolitan Museum, no. 47 1004, found in Ostia, dated about A.D. 200.

[61] Oslo, National Gallery. Vessberg, *op. cit.*, 188–190, pls. XXXII, 2–3 and XXXIII. Levi, Doro, *Records of the Princeton Museum of Art* **3**, 2: 9–18, figs. 1–19, 1944. He calls the material chalk, probably soft limestone, and the position the "usual of the togatus." It certainly is a palliatus. Roman-British portrait of the fourth century.

[62] Benndorf-Schöne, *op. cit.*, no. 102. Mus. no. 290. Found in the Calixtus catacombs. For other examples of busts in the attitude of the palliatus, see appendix to notes 19–20.

[63] Stuart Jones, *Sculptures of the Capitoline Museum*, 38–39, Atrio no. 36*a*, pl. 8, fig. 55 from Phot. German Arch. Inst. Rome, 1934, 290. Early third century.

FIG. 54. Sarcophagus of Annios Katos, Lateran Museum no. 290. From the Calixtus Catacombs. Phot. Moscioni 8273.

that of a woman in the attitude of the palliatus and with the hairdress of the third century. Three full statues of women in a similar draping and pose are in the Musée Alaoui in Tunis.[64] This female type might be called palliata, a woman in the Greek himation, named palla when worn by Roman women.

The "palliata" becomes rather late the counterpart to the "palliatus." Its forerunners are Greek figures like the statuette of Persephone found in the sanctuary of Demeter of Knidos, now in the British Museum (fig. 56).[65] This figure was created in the same period as Sophokles and Aischines, the forerunner of the palliatus. It holds its right hand higher and more to the right than the "Eretria" type. It takes the place of the "Pudicitia" (see figs. 5*b*, 11–12, 28) on tombstones not before the late republican or Augustan period (figs. 25–27, 33). Perhaps the latest example of "Pudicitia" is an early Augustan relief in Aquila. She is surrounded by five men, all in the same type of palliatus.[66] The "Pudicitia" was probably too complicated for the humble artists working on tomb reliefs. The female counterpart was rendered in the most diverse way, in contrast to the male which remained the same through the centuries. Then by and by the palliata was used instead of the Pudicitia for portrait statues, such as those of Livia in the Villa of Mysteries [67] and for Viciria, the mother of Balbus

[64] Musées de l'Algérie et de la Tunisie, **7**: *Catalogue du Musée Alaoui*, 52, pl. 14, no. 39, 1897; **15**: pl. 30, no. 1, pl. 32, no. 3, 1910. See also, the statue from Thasos, Collignon, *Statues funéraires*, 166, fig. 97; a half figure from Thera, *ibid.*, 302, fig. 190. A young girl of the first century A.D. in Egypt: Graindor, Paul, *Bustes et statues portraits d'Égypt Romaine*, Cairo, Univ. Égyptienne, Le Caire no. 11, 117 f. pl. L no. 970.

[65] Smith, A. H., *British Museum catalogue of sculpture* 2: 205, no. 1302. Ashmole, B., *Jour. Hellen. Studies* **71**: 25–26, 1951. Appendix 2, pl. X*c*. *Cp.* similar terracottas, *ibid.*, pl. IX*d* and X*d*.

[66] Romanelli, P., in *Bulletino comunale* **71** (1943–1945), 4–6, figs. 1–3, 1945.

[67] Livia: Maiuri, A., *La Villa dei Misteri*, 223–232, figs. 93–98.

FIG. 55. Sarcophagus of a woman in attitude of palliata. Capitoline Museum, Atrio no. 36a. Phot. German Arch. Inst. Rome, no. 34,290.

FIG. 56. Statue of Persephone. From sanctuary of Demeter of Knidos, British Mus. no. 1302. Courtesy of the Trustees of the British Museum.

(fig. 57),[68] as well as on many tomb reliefs [69] and sarcophagi (see below, figs. 61–62*c*).

[68] Viciria: Mus. no. 6168. Ruesch, A., *Guida del Museo Nazionale di Napoli*, no. 20; New Guide no. 12. Hekler, A., *Bildniskunst der Griechen und Römer*, pl. 205a. *Idem*, *Röm. weibliche Gewandstatuen*, 131.

[69] For Roman tombstones see my appendix to notes 19–20. Examples from Attic tombstones of the Roman period with husband and wife in the same draping and attitude of the palliatus and the palliata: Conze, *op. cit.*, nos. 2076, pl. CCCCLV, and 2118. Rieman, *op. cit.*, 48–50, no. 46, pl. 15.

FIG. 57. Viciria, wife and mother of the Balbi, Herculaneum. Naples, Nat. Mus., no. 6168. Phot. Alinari 11160.

FIG. 58. Christ in a pallium. Fragment of a sarcophagus, from Istanbul, in Berlin. Morey, *Early Christian Art*, frontispiece.

The Christian sarcophagi, like other early Christian art, took over the attitude and draping of the pallium, together with many other features of the contemporary pagan art. Tertullian (*De Pallio* VI) converted to Christianity in the year 190, could aptly praise the pallium as the dress of the Christians and contrast it with the toga. Not only the pallium but also the traditional attitude of the "Eretria" youth is given to Christ on the well-known Berlin sarcophagus (fig. 58),[70] but also to the Apostle Paul on the sarcophagus of Junius Bassus (fig. 59) and to an Apostle on the five arches sarcophagus in S. Francesco at Ravenna (fig. 60).[71]

Thus the tradition persisted through six centuries from the second century B.C. to the fourth century A.D., with forerunners and some late comers extending the life of the type to almost a millennium. Social upheavals, political changes, wars for empire and promises of peace, the fall of ancient gods and the rise of a new religion had no influence on this representation of a man draped in a tightly fitted mantle and with a definite attitude of his hands. What is the meaning of this conservatism?

VIII. THE MEANING OF THE ROMANI PALLIATI

Why is the type of palliatus repeated again and again through the centuries, and why is even the

FIG. 59. Central group of sarcophagus of Junius Bassus. St. Peter's, Rome. From Phot. Anderson, 20454. Negative, Fine Arts Department, Columbia University, S 10,800, 675,10.

[70] For Christ, see the fragment (fig. 58) of the celebrated Berlin sarcophagus from Constantinople: Strygowski, J., *Orient oder Rom*, 58–60, pl. II. Morey, C. R., *The sarcophagus of Claudia Antonia Sabina*, in *Sardis* **5**: 1, 30 and 79, pl. 25, 1924. Wulff, O., *Altchristliche und Byzantische Kunst*, 171, fig. 166. Lawrence, Marion, in *Art Bulletin* **10**: 44, fig. 49, 1927. Gerke, F., *Christus in der spätantiken Plastik*, 105, fig. 76; *idem*, *Die Christlichen Sarkophage*, 227–28 and 232, notes 4–6. He wrongly calls the himation "the cynic mantle." Morey, C. R., *Early Christian art*, frontispiece, 25–6, 60, 107, Princeton, 1942, *ca.* A.D. 400.

[71] For Apostles on sarcophagi, see (our figs. 59–60), Gerke, F., *Der Sarkophag des Junius Bassus*, 1936. Morey, *Early Christian art*, 104 and 278, fig. 110; 134–135, 283, fig. 141. Lawrence, M., in *Art Bulletin* **10**: 40, fig. 45, 1927, and **14**: 128–133, fig. 40; *cp.* figs. 46 and 49, 1932. *Eadem*, *Sarcophagi in Ravenna*, Monograph on Archaeology and Fine Arts, College Art Association, no. 2, 15 f., fig. 25 = our fig. 60; *cf.* 8, fig. 14; 15, fig. 46, 1945. Apostles in book illustrations, see Weitzmann, K., *Roll and Codex*, 201–205, figs. 197, 198, 200, 203, 205; *cp.* 142–143, fig. 130.

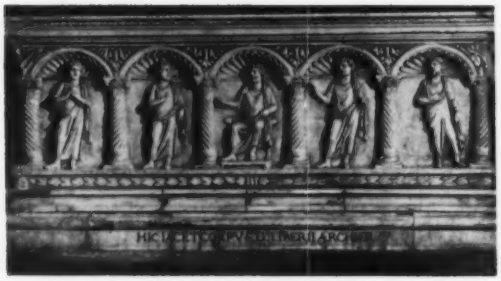


FIG. 60. Five arches sarcophagus, S. Francesco, Ravenna. Marion Lawrence, *Art Bulletin* **10**: fig. 25, 1957. Phot. Alinari, 18075; from negative Fine Arts Department, Columbia University, S 10,850,200.

female counterpart adapted to the same type in the later period? The fact that the monuments so often had empty sockets for fitting portrait heads implies that the figures must have had a definite meaning.

It has been said that these men were orators, based on a description of rhetors by Quintilian, *The Education of the Orator* (*De Institutione oratoria*).[72] This sentence, however, has to be read in connection with the preceding and following sentences (XI, 3, 137–138):

Cultus non est proprius oratoris aliquis, sed magis in oratore conspicitur. Quare sit, ut in omnibus honestis debet esse, splendidus et virilis. . . . Est aliquid in amictu, quod ipsum aliquatenus temporum condicione mutatum est. Nam veteribus nulli sinus, perquam breves post illos fuerunt. Itaque etiam gestu necesse est usos in principiis eos alioquorum brachium, sicut Graecorum, veste continebatur. Sed nos de praesentibus loquimur.

With regard to dress, there is no special garb peculiar to the orator, but his dress comes more under the public eye than that of other men. It should, therefore, be distinguished and manly, as, indeed, it ought to be with all men of position. . . . There are also details of dress which are altered to some extent by successive changes in fashion. The ancients, for example, wore no folds, and their successors wore them very short. Consequently, it follows that in view of the fact that their arms were, like those of the Greeks, covered by the garment, they must have employed a different form of gesture in the exordium from that which is now in use. However, I am speaking of our own days. (Tr. by H. E. Butler.)

There follows (138–149) a detailed description of the toga, how to hold it in order, how Cicero wore it long on account of his varicose veins, etc. Twice in these paragraphs Quintilian contrasts explicitly the toga and the pallium (XI, 3, 139): "togam rotundam esse and apte caesam velint"; and (XI, 3, 143): "togam veteres ad calceos usque demittebant, ut Graeci pallium."

The toga itself should, in my opinion, be round (half elliptical) and cut to fit; the ancients used to let the toga fall to the heels as the Greeks were in the habit of doing with their cloak, the pallium.

It cannot be doubted that, like Cicero, all Roman orators, being public speakers, had to wear the toga, and that the many Romans discussed above could not all be orators.

It has also been said that these men were philosophers or literati, men of the learned class, because several are shown holding book scrolls (figs. 11, 23–24, 36–38, 41, 47, 52, 59), or they are standing beside a case containing such scrolls (see figs. 36–38, 40–41, 43, 45, 47–49). But such a receptacle might also hold documents belonging to a magistrate. The type was certainly used for physicians, as proved by the tomb relief of Clodius Tertius and his father, Aulus Clodius Metrodorus, two physicians from Frascati, now in the Louvre.[73] A tombstone from Crete, now in the British Museum, portrays the physician Aurelius Alexander and is inscribed with his honorary decree.[74] Philosophers, although often holding a scroll, are more often represented seated than standing, because they are supposed to be teaching. Even Sokrates in the later conception of Lysippos, Plato, Epikouros, Metrodoros, Hermachos, Chrysippos, and other Greek philosophers are seated in most representations. The same is true of the seven sages on the sarcophagus in Rome on which one of the philosophers has a portrait head of the deceased.[75] Even Christ is

[72] West, R., *op. cit.*, 41–44, bases on this description the interpretation of the palliatus type as rhetor. However, she calls the mantle a toga. See note 3.

[73] Zadock-Jitters, A., *Ancestral portraiture*, 57 and 75, pl. XVI. Vessberg, *op. cit.*, 201, pl. XLI, 2.

[74] Smith, A. H., *British Museum catalogue of sculpture* **3**: 279, Mus. no. 2243. Phot. Giraudon, no. 34724, 1936. Beside the physician a little boy on a pedestal. Below the physician an honorary decree inscribed on a circular medallion.

[75] See sarcophagus with the seven sages and seven muses in Museo Torlonia. Rodenwaldt, *Arch. Jahrbuch* **51**: 101–103, pl. 5, 1936. Laurenzi, L., *Ritratti Greci*, pl. XLVIII (appendix C): D Plato, E Sokrates, H Chrysippos, I Metrodoros. Bieber, M., *The sculpture of the Hellenistic Age*, 44–47, 55–57, 68–70, figs. 114–115 (Plato); 132–135 (Sokrates); 163–166 (Epikouros); 173–174 (Metrodoros); 238–242 (Chrysippos).

FIG. 61. Sarcophagus from Ephesos. *Mitteilungen des Deutschen Arch. Instituts* **9**: pl. 9, 1952.

seated when represented as a teacher (fig. 59).[76]

Thus the modest citizens on the Attic and Roman tomb-monuments were neither orators nor philosophers. They are bookish people, like the Greeks whom Plautus described, and whom the Romans imitated. While Plautus mocked the disputing Greeks, the fact is that the Romans aspired to Greek culture and Greek education. Both Greek and Roman figures were often shown having scrolls in their hands or boxes with scrolls standing beside them, just as Plautus described the Graeci palliati. The scrolls tell us that their bearers were educated men, who could be of any profession. The pallium, mentioned as the Greek dress by Plautus, was also adopted by the Romans, when they came in touch with the Greek culture. Even Cato, originally a violent enemy of the effeminate Greek achievements, became imbued with them, learned Greek and took the pallium, as Tertullian tells us (*De Pallio* III, 7):

> . . . hoc pallium etsi Graecum magis, sed lingua iam Latinum est. Cum voce vestis intravit. Atque adeo ipse Graecos praeter urbem censebat, litterarum eorum vocemque senex iam eruditus, idem Cato iuridicinae suae in tempore humerum exertus, haud minus palliato habitu Graecis favit.
>
> . . . Although this pallium is more Greek, it is almost Latin. With the word came the dress. Even the old Cato, who censored the Greeks, studied their literature and language and favored the Greek pallium.

The unassuming and restrained draping of so many male palliati and by and by also of the female palliata expressed piety and modesty both for men and women. Therefore, it became the prescribed form for funerary, commemorative, and honorary monuments. Men sacrificing, pouring libations, or fulfilling other religious duties had always to be fully clothed. The toga, prescribed by Augustus for public appearance could meet this requirement, but it was obligatory only for members of the imperial house and high officials. For all others the pallium-himation did just as well.

A sarcophagus from Ephesos exemplifies the use of both these simple types of the palliatus and the palliata for religious purposes (fig. 61).[77] A woman in a palla wrapped tightly around her and a palliatus in the same draping stand together before Pluto and Persephone. Hermes has led them there past the three Fates at the other side. Another palliata stands alone before Pluto and Persephone on a cover used for the Endymion sarcophagus in the Capitoline Museum (fig. 62).[78]

[76] For Christ seated, holding a scroll as a teacher of divine wisdom, see: Kollwitz, J., *Das Christusbild des dritten Jahrhunderts* **12**: 20–22, pls. 3, 6, 7, 1953.

[77] Keil, Osterr. Jh., **17**: 135–144, figs. 78–79, 1914. Curtius, L., *Mitteilungen des deutschen Instituts* **4**: 28–29, pl. 9, 1952. *Cf.* Egger, R., Zwei oberitalienische Mystensarkophage, *ibid.*, 35, pl. 17.

[78] Jones, Stuart, *Sculptures of the Capitoline Museum*, 314–315, Stanza del Fauno 3b., pl. 78. He assumes that the veiled woman represents the woman's soul. Altmann, *Architektur und Ornamentik der antiken Sarkophage*, 98, fig. 16.

a

b

d

c

e

f

FIG. 62. Cover of the Endymion Sarcophagus in the Capitoline Museum, Stanza del Fauno 3*b*. *a*. Phot. Anderson, 1766; *b–f*. Phots. Renato Sansaini. *a*. The entire cover. *b*. The couple kneels to the Fates. *c*. The deceased in the attitude of the palliata. *d*. Pluto and Persephone. *e*. Mercury. *f*. The couple in everyday dress on a couch.

At the right she is seen seated with her husband on a couch, both in everyday dress (fig. 62*f*). To the left she and her husband are kneeling before the three Fates, asking for mercy, so that they would not be separated, again in ordinary dress (fig. 62*b*). But when Hermes (fig. 62*e*) has taken the wife to the Netherworld, she stands before Pluto and Persephone near to their altar, fully wrapped, in the ritual attitude, in a pose of deep reverence (figs. 62*c–d*). The dead must approach the gods with humbleness and humility, and the drapery reflects this attitude.

Besides religious, there were two other reasons why this type was retained through Hellenistic, Roman, and early Christian times. The himation was a pleasant dress, and the Hellenistic form in which it was represented must have pleased the Romans. When using a general model, be it for a Doric column or a statue, the Greeks always changed details in each new undertaking, trying to improve upon the preceding one. In contrast, the Romans adhered to stereotyped forms, as the Etruscans and the Egyptians did before them. It did not matter whether the object was a dress, a tomb relief, or a statue. It was reported monotonously down to the last detail. The toga had to have its long end hanging between the feet, its deeply curved overfold (sinus) and its knot (umbo) before the breast. Truly, the toga was a uniform as compared to the Greek dress with its unending possibilities of draping.[79] But when the Romans chose figures of the "Pudicitia" or other Greek types for their portrait statues, they repeated them again and again in the same way. Thus we might surmise that the artistic form of the palliatus must have pleased them, for they kept it through many centuries with only minor variations. Of course, the stylistic execution was different in different periods.

Another reason is that the himation was a pleasant dress not only for art, but for life. We see it so frequently in art because it was a popular costume for everyday wear (see my models figs. 2*b* and 51). It was applied to portrait statues and funeral monuments with good reason, for it represented the people as they actually were. The himation was much more convenient than the toga, the folds of which had to be arranged in a prescribed manner and were not supposed to be changed even by movement. Although the high officials, senators, and the emperors were supposed to wear the toga in public, they sometimes neglected the formal attire in favor of the pallium. Scipio Africanus, when in Sicily, discarded his official dress and walked about in pallium and sandals (Livy, XXIX, 19, 12). Senator Rabirius, when in Alexandria, wore the pallium (Cicero, *Pro Rabirio* IX, 27). Verres wore a dark tunic and an himation when he supervised the resetting of the emblemata which he had stolen (Cicero, *Actio in C. Verrem* II, 4, 54 and V, 33, 52). The Emperor Tiberius adopted the pallium and sandals while in Rhodes (Suetonius, *Tiberius* 13). Hadrian donned the pallium at imperial banquets, and when he was agonothetes at the Dionysia (*Script. Hist. Aug., Hadrian* XXII, 4–5. Dio 69, 16, 1). This is in agreement with his statue from Cyrene (fig. 38). Septimius Severus came to a banquet of Marcus Aurelius, who was then emperor, wearing a pallium. But, as this was not considered a proper dinner garment, he was given a toga from the imperial wardrobe (Suetonius, *Severus* I, 7). In the last quarter of the third century, the emperors Probus and Carinus occasionally wore pallia of scarlet and purple, interwoven with gold (*Script. Hist. Aug., Probus* IV, 5 and *Carinus* II, 5). Literary and monumental evidence both point to the fact that the himation-pallium was widely used by the Romans from the republican to the late Roman period, in everyday life as well as in art.

We may draw an inference from the comparison of the monumental with the literary evidence. The high personalities like Scipio, Rabirius, the Emperor Tiberius took the pallium when in Greek lands, Magna Graecia or the Greek islands. Septimius came from North Africa, which also had originally a Greek civilization. Thus they adopted the pallium when they came in touch with the Greek culture. The Emperors Hadrian and Julian the Apostate, who wore the pallium, were outstanding Graecophiles. Yet it seems to me that most people who are represented wearing the pallium in art, are members of the upper middle class, the one in every society who cultivates art and literature. To the Romans that meant Greek art and Greek literature.

As the Christians came mainly from the middle class, or from Greek slaves, at least in the beginning, it is understandable that they also adopted the pallium. Tertullian, who became a Christian

[79] *Cp.* Wilson, L., *The Roman toga*, and Bieber, M., *Griechische Kleidung*, 21–24, 54–77, pls. XXIII–XLII. For the characteristics of the toga, *cp.* Wilson, L., *The clothing of the Romans*, 46 and 78–83.

in 190 and died in 230, described the pallium as the dress of cultivated men:

De Pallio VI, 2:

De meo vestiuntur et primus informator litterarum et primus enodator vocis et primus numerorum harenarius et grammaticus et rhetor et sophista et medicus et poeta et qui musicam pulsat et qui stellarum coniectat et qui volaticam spectat. Omnis liberalitas studiorum quattuor meis angulis tegitur.

Sed ista pallium loquitur. . . . Gaude pallium et exsultat! Melior iam te philosophia dignata est quo Christianum vestire coepisti.

In me are dressed the teacher of literature and language and the mathematician and grammarian and rhetor and sophist, and the physician and the poet and musician and astronomer and the observer of birds. All liberal arts are contained in my four corners.

But this is spoken by the pallium. . . . Rejoice, o pallium, and exult! You are now even more dignified than philosophy, because you have begun to dress the Christian.

No wonder that the type lived on for Christ and the Apostles (see figs. 58–60) in the early Christian and even into mediaeval times.[80]

I suggest calling the type "Romani Palliati" as a parallel to Plautus' "Graeci Palliati."

APPENDIX TO NOTES 19–20

Romani Palliati in Tomb Reliefs

a. FULL FIGURES

Rome:

Museo nuovo. Mustilli, *Il Museo Mussolini*, p. 102, no. 9, pl. LVI, fig. 255. M. Bieber, *Entwicklungsgeschichte der griechischen Tracht*, 52, pl. 49, 1. O. Vessberg, *Studien zur Kunstgeschichte der römischen Republik*, 186 ff., 211 f., 249 f., pls. XXVII–XXVIII. With "Pudicitia." About 80–50 B.C. Phot. Inst. Rome 29.172. Our fig. 12.

Museo Nazionale Romano, Garden Court outside the Museum. Vessberg, *op. cit.*, 197, pl. XXXVI, 2. Very elongated figure. *Ca.* 30 B.C. *Ibid.*, No. 105054. Two men in same attitude. Lower part broken. Phot. Inst. Rome, 37.812–814. Our fig. 23. *Ibid.*, three torsos at the back wall of the Garden Court. Marble. One much worn.

Vatican. Giardino della Pigna No. 91. W. Amelung, *Skulpturen des Vatikanischen Museums* **1**: 845, pl. 99. Life size. Travertine.

Galleria in Villa Borghese. Twice to the sides of the "Pudicitia." L. Wilson, *The clothing of the Romans*, 152, pl. LXXXVIII, fig. 96. Phot. Alinari 27481. Our fig. 11.

Villa Doria—Pamphili. Man with wife and daughter. Phot. Institute Rome, no. 21719.

Villa Medici. Head of Septimius Severus not belonging. Cagiano de Azevedo, *Le Antichità della Villa Medici*, 45, nos. 17–20, pl. XII, 18.

Palatine, opposite the entrance from Palace of Tiberius to the cryptoporticus leading to the House of Livia.

Accademia di San Luca. Garden Court in Via della Stamperia. Torso. Nearby a "Pudicitia."

From the Tomb of Eurysaces? Outside Porta Maggiore. Early Augustan. Vessberg, *op. cit.*, 195, pl. XXXIV, 3. Phot. Institute Rome no. 1933, 749. Our fig. 31.

From the Tomb of the Scipios. Phot. Inst. Rome, 37.948.

From the Via Appia. Three headless torsos. Phot. Inst. Rome, 34.1917–1920. Our figs. *22a–d*.

Capua., Museo Campana. Two men in same attitude. Phot. Inst. Rome, 4685. Our fig. 24.

Aquila. Five men in same type. Woman "Pudicitia." P. Romanelli, *Bulletino comunale* **71** (1943), 1945, pp. 4 ff., figs. 1–3. Early Augustan.

Soriano nel Cimino, Villa Chigi. Phot. German Inst. Rome, 30.659.

Trieste. Phot. German Inst. Rome, 37,660. Our fig. 32.

London, British Museum, no. 2243, from Crete. The physician Aurelius Alexander standing above a roundel containing an honorary inscription. Next to him little boy on pedestal. *British Museum catalogue of Sculpture* **3**: 279, fig. 38.

Athens, see note 7.

Asia Minor. See note 8.

Southern Russia. See Watzinger-Kohte, Griechische *Grabreliefs aus Südrussland* (1909), pl. XVII, nos. 243, 254, 257; XXIV, nos. 353, 355; XXXIII, no. 468; XXXIV, no. 486. Blavatsky, *The art of the northern coast of the Black Sea*, 22, fig. 8, 1947. Man and woman in same attitude.

Palmyra, now Kopenhagen. Wulff, *Altchristliche und Byzantinische Kunst* **1**: 133 f., fig. 123. M. Lawrence, City gate sarcophagi, *Art Bulletin* **10**: 44, fig. 48 (third century A.D.), 1927.

b. BUSTS

Rome:

Museo Nazionale Romano. Garden Court outside the Museum, no. 1255. Aristo, Loccius and wife Agate in same attitude. Agate with mantle over head. *Ibid.*, no. 39475, Wife "Pudicitia." Both old. Aelius and his wife Athena. Three-quarter figures. Both liberti. Athena same type, but head veiled. Tuff. Hairdress of Octavia. Vessberg, *op. cit.*, 186 f., note 1. Two men with wife and young son (in full figure). Vessberg, *op. cit.*, 185 f., pl. XXIX, 1. Staglia and Valerius. Wife "Pudicitia." *Ibid.*, Chiostro di Michelangelo, No. 125655. Old man and woman in same type. Between them young man holding fruits. Dated 40–30 B.C. Arndt-Amelung, *Einzelaufnahmen ant. Skulpt.*, no. 3243. *Ibid.*, in garden of the Chiostro, on a peperino

[80] Christ as a palliatus standing before Pilate, from a codex purpureus, dated A.D. about 500 or later. All other figures have the Byzantine robes. Munoz, *Il codice purpureo di Rossano* **5**, pl. XIV, 1907. Wulff, *op. cit.*, 302–303, fig. 282.

wall. Tombstone of the family of the Fonteii. Three men and one woman in same attitude.

Capitoline Museum. Rupilius and wife and son. All three same type, wife mantle over head. Dated 70–60 B.C. Stuart Jones, *Sculptures of the Capitoline Mus.*, 122, No. 48*a*, pl. 21. R. West, *Römische Porträtplastik*, **1**: 55, pl. XII, fig. 46. Vessberg, *op. cit.*, 187 f, pl. XXVI, 2–3. Old man dead, flanked by two living members of the family (Inscr. VIV). Stuart Jones, *op. cit.*, 64 f. Stanza Terrena a dritta, No. 5, pl. 12.

Museo Nuovo. Blaesius and Blaesia. Both same attitude, Blaesia mantle over head. In Giardino of former Palace Caffarelli. No. 2279. Dated about 75 B.C. Mustilli, *Il Museo Mussolini*, p. 176, no. 60. Vessberg, *op. cit.*, p. 183 f., pl. XXV, 2. Man with wife. *Ibid.*, no. 2282. Dated 90–60 B.C. Vessberg *op. cit.*, 184 f., pl. XXVI, 1. *Ibid.*, 6 busts, of which four men, same attitude. One woman "Pudicitia," the other hairknot. Dated 30–15 B.C. Vessberg, *op. cit.*, 200, pl. XL, 2.

Museo del Laterano. Tomb relief of the Furii, no. 742. The two men same attitude, the three women differentiated. They have hairknot. Dated 35–30 B.C. Benndorf-Schöne, *Bildwerke des Lateranensischen Museums*, no. 467. West, *op. cit.*, **1**: 141 f., pl. XXV, fig. 149. Vessberg, *op. cit.*, 197 f., 232, pl. XXXIX, 2. Phot. Anderson, no. 24229 = our fig. 30. Servilius with wife and son, no. 723. Period of Augustus. Benndorf-Schöne, *op. cit.*, no. 23. Vessberg, *op. cit.*, 192 and 201. Altmann, *Die römischen Grabaltäre*, fig. 158. Epictes with wife and daughter. Travertine. About 50–40 B.C. Benndorf-Schöne, no. 454. Arndt-Amelung, *E. A.*, 2214. Vessberg, *op. cit.*, 190, pl. XXXI, 2. Fragmentary relief. *Ca.* 50 B.C. Vessberg, 190, pl. XXXII, 1. Chiostro of San Giovanni in Laterano. Family of the Cavii. Grandfather, father and son in the same attitude. Vessberg, *op. cit.*, 186 f., pl. XXXI, 3. Phot. German Institute Rome, 6537.

Vatican. Museo Chiaramonti, no. 9. P. Aelius Verus. Two men in same type, at the side of woman and child. Period of Augustus. W. Amelung, *Skulpt. Vat. Mus.* **1**: 321 f. Chiaramonti 13*a*, pl. 32. Museo Chiaramonti, no. 21. Vibius with wife in "Pudicitia" type and boy. Julio-Claudian period. Amelung, *op. cit.*, **1**: 348. Chiaramonti, no. 60E, pl. 36. Hekler, *Bildniskunst*, pl. 134. West, *op. cit.*, 53 ff., pl. XXI, fig. 45. Vessberg, *op. cit.*, 201 f., pl. XLI,3. Charbonneaux, *L'Art au Siècle d'Auguste*, fig. 18.

Museo Torlonia. Album del Museo Torlonia, pl. CXVII, no. 456. *Ibid.*, pl. CXVIII, no. 460.

Villa Celimontana. Ca. 70–60 B.C. Vessberg, *op. cit.*, 187 f., pl. XXX, 1–2.

Villa Colonna. Tomb relief of the Manlii. About 45–40 B.C. One young, one old. Vessberg, *op. cit.*, 198 ff., pl. XXXVIII, 1. Phot. German Arch. Inst., 1933, 1717.

Villa Doria-Pamphili. Tomb relief of the Visellii. Two men in the same attitude. Vessberg, *op. cit.*, 197. pl. XXXIX, 1.

Via Appia. Rabirius. *Ca.* 50–20 B.C. Vessberg, *op. cit.*, 191, pl. XXXI, 1.

Avellino, Treviso, Ascoli Piceno, and *Beneventum.* See note 20 and our figs. 25–29.

Pozzuoli. Amphitheater. Claudian period. Phot. German Inst. Rome, No. 37,643. Our fig. 33.

Boston, Museum of Fine Arts. About 45–30 B.C. The younger Gessius, with father and wife. Found in or near Rome. Vessberg, *op. cit.*, 190 f., pl. XXXV.

Paris, Louvre. Found in Frascati. Two physicians, Clodius Tertius and his father Aulus Clodius Metrodorus. Zadock-Jitters, *Ancestral Portraiture*, 57, 75, pl. XVI, B. Vessberg, *op. cit.*, 201, pl. XLI, 2.

London, British Museum. Ampudius. About 30–20 B.C. A. H. Smith, *Jour. Roman Studies* **8**: 179, pl. II, 1918.

Berlin, Altes Museum. Aiedius. Augustan period. Blümel, *Römische Bildnesse*, 3 f. R 7, pl. 14. West, *op. cit.*, 53 and 99, pl. XII, fig. 44. Vessberg, *op. cit.*, 203 f., pl. XLIII, 2.

Copenhagen, Glyptotek Ny Carlsberg. Septumius. Found in Vulci. *Ca.* 40 B.C. F. Poulsen, *Cat. of ancient sculpture in the Ny Carlsberg Glyptotek*, no. 556. *Billedtavler*, pl. XXXXIV, 1951. Arndt, *Griech. und röm. Porträts*, pl. 251. West, *op. cit.*, **1**: 50, pl. X, fig. 34. Vessberg, *op. cit.*, 202 f., pl. XLII, 1. *Ibid.* Formerly Landsdowne Collection. Time of Tiberius or Claudian period. F. Poulsen, *op. cit.*, no. 591*b*. 2. *Tillaeg til Billedtavler* (second plate supplement), pl. IX. Arndt-Amelung, *Einzelaufnahmen ant. Skulpt.*, nos. 3049–51.

For busts in the round see notes 59–63.

For figures on sarcophagi see notes 55–58, and 70–71.

IMPROVISATION IN THE MIDDLE ENGLISH ROMANCE

ALBERT C. BAUGH

Felix E. Schelling Memorial Professor of English, University of Pennsylvania

(*Read November 13, 1958*)

I

In trying to recapture the conditions under which the Greek epic and the Old French epopee were created and recited—conditions under which the conveyance of long narrative poems was oral—much use has been made in recent years of the presumably parallel conditions which exist in Jugoslavia and elsewhere. In Jugoslavia and other parts of the world long heroic poems are still recited or sung to the accompaniment of a musical instrument by singers who in many instances are unable to read and to an audience which is equally dependent upon the ear. While such conditions exist in parts of Russia, in Turkish areas, in Siberia, Crete, and elsewhere,[1] those in Jugoslavia have been the most studied and are the best known. A comparatively early study of Jugoslavic oral poetry was that of Mathias Murko,[2] but the most extensive studies have been those of Milman Parry and A. B. Lord.[3] The Jugoslavic evidence has been used especially for the elucidation of the Homeric poems, but recently Jean Rychner has used it to propound a theory of the composition of the early chansons de geste.[4]

There are several respects in which the Greek epic, the Old French chansons de geste, and the poems sung by Jugoslav *guslars* appear to be comparable. In subject matter they are essentially heroic narratives. They are sung to the accompaniment of a musical instrument. In Greece it appears to have been a lyre; when the celebrated minstrel Demodocus in the eighth book of the *Odyssey* relates the heroic deeds of the Achaeans he sings them to the accompaniment of the lyre. In the chansons de geste the instrument most frequently mentioned is the *viele* and the jongleur is said to "chanter et vieler ensemble."[5] The Jugoslav instrument is a *gusle*, a primitive sort of violin of one string played with a crude bow. Most important, however, is the fact that all these heroic poems are at times very long. Milman Parry recorded the song of one *guslar* which ran to 13,331 lines,[6] which is longer by a thousand lines than the *Odyssey*, and a Kara-Kirghiz bard, shortly before his death in 1930 told the story of the Mohammedan hero, Manas, in some 40,000 lines.[7]

Now the thing to be noted in the practice of all singers of heroic songs in Jugoslavia and in other parts of Europe and Asia where their art has been studied is that the singing of these poems involves a considerable element of improvisation. For one thing, singers often sing the same story or instalment in versions of quite different length—in a version that takes twenty minutes or one that takes an hour or more. It is the same story but in different degrees of fullness. The shortening is not accomplished by skipping incidents or omitting passages. It is general condensation and each is a separate version. Yet it is not a matter of knowing a longer and a shorter song, both of which are in the singer's memory and from which he chooses one or the other as the time at his disposal requires. He can expand or contract at will and do it in the very act of singing. The result is that no two performances are exactly alike. Murko reports that on one occasion he had a singer dictate his song before singing it. The intention was to record any variants while listening to the song. Although only a few minutes intervened, the version which the singer sang was

[1] See Radlov, V. V., *Proben der Volkslitteratur der türkischen Stämme Süd-Siberiens und der dsungarischen Steppe*, St. Petersburg, 1866–1907; *Trans. Amer. Philol. Assoc.* **82**: 84; C. M. Bowra, *Heroic poetry*, chap. 6, London, 1952.

[2] *La poésie populaire épique en Yougoslavie au début du xx^e siècle*, Paris, 1929.

[3] Most important in this connection is the article of Milman Parry, Studies in the epic technique of oral verse-making: I. Homer and Homeric style, *Harvard Stud. in Class. Philol.* **41**: 73–147, 1930.

[4] *La chanson de geste: essai sur l'art épique des jongleurs*, Geneva and Lille, 1955.

[5] *Le roman de la violette*, line 1402, describing Gerard the hero disguised as a jongleur and singing a passage from *Aliscans*.

[6] *Trans. Amer. Philol. Assoc.* **52**: 42.

[7] *Ibid.*, 86.

so different from the dictated copy that it was quite impossible to record the differences as variants. This is the situation which the editor of a medieval text sometimes finds himself in when he is dealing with more than one manuscript. Moreover, the theory of improvisation is supported by the testimony of singers themselves. A. B. Lord in his essay, "Composition by Theme in Homer and Southslavic Epos"[8] says:

> A Yugoslav singer told me last year that when he learned a new song he made no attempt at word-for-word memorization but learned only the "plan" of the song, which he explained as "the arrangement of the events." This plan he then proceeded to fill in with the themes which he already knew.

In other words he partly improvised.

It is bound to cause some surprise to be told that an illiterate singer is able to compose his verses as he goes along. The secret of this apparently extraordinary ability lies in the fact that improvisation is not a sheer act of creative composition, but the adroit utilization and recombination of elements in a well-stocked memory. These elements are of two kinds: formulas and themes.

A *formula* in the definition of Milman Parry, based primarily on the Homeric poems but equally applicable to all epic poetry, is "a group of words which is regularly employed under the same metrical conditions to express a given essential idea." Everyone will recall "the blue-eyed goddess Athene" (θεὰ γλαυκῶπις Ἀθήνη) since the phrase occurs more than fifty times, or τὸν δ' αὖτε προσέειπε (however, he said to him), which occurs nearly ninety times. There is no reason why the poet should so regularly use these stock expressions rather than seek greater variety except that they furnished ready-made phrases of a certain metrical pattern easily fitted into the line. In the same way, as Parry says, this is

> likewise true of the epithet ἄναξ ἀνδρῶν, "lord of men," which is used not only 48 times elsewhere in the *Iliad* and the *Odyssey* of Agamemnon, but also of Anchises, Aeneas, Augeias, Euphetes, and Eumelos, none of whom is any more a king of men than is any other of the chief heroes, but all of them have names of the same metrical value as that of Agamemnon.[9]

Rychner finds a similar use of formulas in the early chansons de geste: for example in the nine chansons which he studied he found 39 instances of

> Le destrier broiche
> Le cheval broche
> Les destriers broichent
> Son cheval brochet, etc.

at the beginning of the line, which was then completed with other formulas such as

> des esperons tranchanz
> des esperuns d'or fin
> des esperons burnis
> des esperons forbiz
> des esperons d'or mier
> des esperons des pies
> des oriez esperuns
> des aguz esperuns
> le frein li abandunet, etc.

according to the requirement of the assonance. And he cites many other formulas similarly used.

The second of the traditional elements which the singer makes use of in telling his story is the *theme*. In his essay "Composition by Theme in Homer and Southslavic Epos," previously referred to, A. B. Lord defines the theme as "a recurrent element of narration or description in traditional oral poetry," and he adds:

> It is not restricted, as is the formula, by metrical considerations; hence, it should not be limited to exact word-for-word repetition. . . . Regular use, or repetition, is as much a part of the definition of the theme as it is of the definition of the formula, but the repetition need not be exact.

As examples of such recurring elements we may mention such stock situations as the boastful speech or taunt or challenge, the arming of the hero or the description of his arms, his mounting and spurring his horse to the fight, the description of the opposing army, description of a mêlée, the individual combat, injuries before, during, or after the fight, regret for a hero killed or seriously injured, prayers, greetings and leave takings, feasts, the dubbing of a new knight, and so on. In the chanson de geste Rychner finds that the theme often constitutes a laisse and that an entire laisse is sometimes repeated with only slight differences. Thus in the *Chanson de Guillaume* a sequence of three laisses totaling 42 lines describes the death of Girard and immediately afterwards the same three laisses are used with some omissions and additions to describe the death of Guichard.[10] Likewise in describing the arming of a knight, a theme that naturally occurs frequently, a stock sequence of ideas with a

[8] *Trans. Amer. Philol. Assoc.* **82**: 73–74, 1951.
[9] *Op. cit.*, 123.
[10] Rychner, 83–84.

number of identical lines is used for Thibaut, Girard, and Guillaume at intervals of 400 to 900 lines:

Dunc li vestent une broine mult bele (e cler),
E un vert healme li lacent en la teste;
Dunc ceint s'espee, le brant burni vers terre,
E une grant targe tint par manvele;
Espé trenchant out en sa main destre,
E blanche enseigne li lacent tresque a tere.
Dunc munte Tidbald par sun estriu senestre.
(133–139).

Dunc li vestirent une broigne mult bele,
E un vert healme li lacent en la teste.
Willame li ceinst l'espee al costé senestre;
Une grant targe prist par la manvele;
Cheval out bon, des meillurs de la terre.
Puis muntad Girard par sun estriu senestre.
(1075–1080).

Dunc li vestirent une broine mult bele,
E un vert healme li lacent en la teste;
Sa espee out ceinte, le brant burni vers terre,
Une grant targe i tint par manevele;
Espé trenchante out en sun poig destre.
(1498–1502).

Indeed the formula is even used for the collective arming of the Saracens:

Cent mile furent de la pute geste;
Il n'i out celui de blanc halberc ne se veste,
E de Saraguce verz healmes en lur testes,
D'or les fruntels e les flurs e les esses,
Espees ceintes, les branz burniz vers terre,
Les bons escuz tindrent as manveles,
Espees trenchanz e darz as poinz destres,
Chevals coranz d'Arabe suz lur seles.[11] (220–227).

Rychner's thesis has not gone unchallenged, but his evidence is arresting and, although too new to have been subjected to extensive examination, has been supported by Adrien Bonjour in a sympathetic critique in a recent number of *Romania*.[12] The hypothesis of formulaic elements and oral improvisation has been proposed by Professor Magoun for the *Beowulf*.[13] I do not mean to imply that I consider the hypothesis proved either for the Homeric poems, the chansons de geste, or the Anglo-Saxon epic. In fact, I do not see that it is susceptible of proof in the absence of such direct evidence as we have for the contemporary practice of singers in Jugoslavia and other comparable areas. But since the practice of improvisation can be observed in our own day, since it results in the necessary use of formulaic elements, and since similar formulaic elements are found extensively in the epic poems of Greece, medieval France, and Anglo-Saxon England, the possibility cannot be denied that here also they have resulted from comparable circumstances of oral presentation. It would seem reasonable, therefore, to ask whether the Middle English romances as they have come down to us are to any extent the product of similar practices.[13a]

II

Every one knows that the Middle English romances are honeycombed with stock phrases and verbal clichés, often trite and at times seemingly forced. A number of extensive collections of these commonplaces have been made.[14] They have generally been cited as parallels between different romances showing the conventional character of the ideas and expressions of the poets who wrote them. The attitude of modern critics toward this feature of romance style is on the whole unfavorable, although there have been apologists. Saintsbury thought "that repetition, stock phrases, identity of scheme and

[11] Rychner, 133–134.

[12] **78**: 243–255, 1957.

[13] Oral-formulaic character of Anglo-Saxon narrative poetry, *Speculum* **28**: 446–467, 1953.

[13a] The present study is confined to those romances which make up the so-called Matter of England: *King Horn*, *Havelok*, *Beves of Hampton*, *Guy of Warwick*, *Richard the Lion-Hearted*, and *Athelston*.

[14] So far as I know, the first to gather a large number of them together was Zupitza, in the notes to his edition of the Cambridge Univ. Library MS of *Guy of Warwick* (Early English Text Soc., Extra Ser. xxv–xxvi, 1875–1876). In his inaugural dissertation of 1879 on *Sir Orfeo* Oscar Zielke assembled and roughly classified a representative selection of them. The material was incorporated in his edition of the poem the following year. Many were again cited by Kölbing in his editions of *Sir Tristrem* (Heilbronn, 1882), *Amis and Amiloun* (Heilbronn, 1884), *Beves of Hampton*, EETSES, xlvi, xlviii, lxv), *Athelston* (*Englische Studien* **13**: 331–414, 1889), *Ipomedon* (Breslau, 1889), and *Arthur and Merlin* (Leipzig, 1890). The third volume of the *Beves* contains an essay by Carl Schmirgel, Typical expressions and repetitions in Sir Beves of Hampton, originally printed at Breslau in 1886. Repeated phrases of a more special type are catalogued and discussed by J. S. P. Tatlock in his: Epic formulas, especially in Laȝamon, *Publ. Mod. Lang. Assoc.* **38**: 494–529, 1923, a few of which are found in later romances. A. M. Trounce offers a classification of the uses to which they were put in the tail-rime romances (*Medium Ævum* **1**: 168–182, 1932). Miss Ruth Crosby mentions many of the shorter and commoner clichés in her article: Oral delivery in the Middle Ages, *Speculum* **11**: 88–110, 1936.

form, which are apt to be felt as disagreeable in reading, are far less irksome, and even have a certain attraction, in matter orally delivered.[15] Tatlock suggested as a compensating virtue the fact that, "The well-tried phrase for what is usual leaves the full sharpness of the attention for what is fresh,"[16] and he elsewhere says, speaking of the epic formulas in Layamon, "Doubtless his auditors too found a pleasure in the repetition, as children feel in *The Three Bears* and *The House that Jack Built*, or even as we feel in recognizing a recurring *motif* in music."[17] There is doubtless an element of truth in all these observations, and if the use of conventional phrases had not so often made them into mere tags—a device for filling out the line or meeting the exigencies of rime—one might consider them a conscious artistic device, but as Tatlock observes, "they are apt to be brief, parenthetical and unessential, making little or no contribution to the narrative . . . a general stock of insignificant, shop-worn counters, the profusion of which suggests helplessness."

But are they mere clichés? We should perhaps not answer this question too hastily. By far the largest number occur at the end of a line, where their convenience for purposes of rime is obvious. Each fits into a given metrical situation—the last four syllables of a line, or the last three, or the last five. In other words, they meet the requirements of the formula. A few examples will make this apparent:

Þe king þar of	was glad & bliþe	Beves	529
Þe king þar of	was glad & bliþe		905
And Beues	was glad & bliþe		2497
Þo was Beues	glad & bliþe		3471
Of whom þat he	was glad & bliþe	Guy	1924*
As he þat	was glad & bliþe		6394
Euere he	was glad and bliþe	Hav.	947
And maden hem	[ful] glad and bliþe		1245
Bes of him	ful glad and bliþe		2246

Other phrases such as *and was ful bliþe*, *so was he bliþe*, or, where a syllable less is needed, *swiþe bliþe* satisfy the metrical requirements in other thought contexts.

Adverbial phrases are particularly likely to become formulas because they may be so freely added to complete an idea:

Þai gonne schete	be ech a side	Beves	882
Wiþ gret ioie	be ech a side		2162
Þe cri aros	be ech a side		4438
& mi lond destrud	in ich a side	Guy	2878
Þe messangers	by ylke a syde	RLH	147
And slowȝ dounryȝt	on ylke a syde		5236
He sente aboute	on ylke a syde		6594
Out at is mouþ	in aiþer side	Beues	809
Gret slauȝter worþ	in eiþer side		4130
To hire godes þai bede	in eiþer side		4144
Mani on þer dyed	in aiþer side	Guy	3025
Þar was ioie	be eueri side	Beves	3962
& þer-fore	on euerich a side	Guy	969
Þus þai leide on	in boþe side	Beves	1755
He was be-set	in boþe side		4399
And smot him þourȝ	out boþe side		1002
And hys hoostes	on boþe syde	RLH	1654

Here the variants *ech a*, *ylke a*, *eiþer*, *boþë* are metrically equivalent, while a different requirement is met by *be his side:*.

& fiftene Sarasins	be is side	Beves	588
& dede lede Arondel	be is side		1508
A swerd a tok	be his side		1599
& pyk and skrippe	be is side		2241
Ten þousand Sarezynes	by here side	RLH	7120

With a substitution of different prepositions the rime-word *cité* may enter into a variety of almost identical formulas:

At Warwike	in þat cite	Guy	187
& al þat weren	in þat cite		2260
Worþschipliche	in þat cite		5040
Þe noise aros	in þat cite		6436
Off syluyr & gold	in þat cyte	RLH	4603
Noyse & cri he herd	in þat cite	Guy	2897
We wil abide	in þis cite		5065
& gon oȝain	to þat cite		5372
Alle sori in	to þat cite		6488
Þat hij come	to þat cite		6514
Wel riȝt he ȝede	to þat cite		7023
þus þey come	to that cytee	RLH	4204
Þai wenten out	of þat cite	Guy	2349
Alle þat weren	of þat cite		3867
& þe citiseins	of þat cite		5503
Þo ich went out	of þat cite		5972
By þat on syde	off þat cytee	RLH	5406
At the sege	of þat cytee		5424

Metrically equivalent but suiting a different construction is *þat riche cite:*

[15] Saintsbury, George, *The flourishing of romance and the rise of allegory*, 49, New York, 1897; cited by Miss Crosby, *op. cit.*, 104.

[16] *Op. cit.*, 528–529.

[17] *Ibid.*, 513. The most enthusiastic defense of the device, in its extensive use in the tail-rime romances, is that of A. M. Trounce, in the article cited in note 14, above.

Toward Acres	þat riche cete	RLH	634
Ierusalem,	þat ryche cyte		5898
To Arascoun,	þat gode cite	Guy	1976*
To Espire	þat riche cite		2794
& feffe þe wiþ mani	a riche cite		3352
Mombraunt is	a riche cite	Beves	2045
To London, to	þat riche cite		4536

Contre may be substituted for *cite:*

Boute seue mile	of þat contrei	Beves	1696
Ne com þer non	in þat contre		2247
Boute fif mile	of þat contray		3686
Ich was y-born	in þat cuntre	Guy	938
Bot al þe folk	of þat cuntre		965
& al þe folk	of þat cuntre		1113
Ich was y-born	in þat cuntre		1746

Other monosyllabic adjectives or pronouns may replace *þat* without disturbing the metrical value of the formula:

Hom te wende	to his contre	Beves	4023
What they dyde	in his countre	RLH	712
As þou hast folk	in þis cuntree		6463
A kniȝt icham	of fer cuntre	Guy	1635

There are, of course, many more.

One of the most useful formulas not so plainly used to fill out the line is the type which ends in a past participle. The verbs most frequently used are *y-brouȝt* and *y-go(n)*, although the latter has substitutes available (*fare*, *ride*, etc.):

þat neye he was	to deþ y-brouȝt	Guy	246
Ich wold ich were	to deþ y-brouȝt		336
þat he is	to deþ y-brouȝt		1456
Hou mine men ben	to deþ y-brouȝt		3530
Who haþ mi lyoun	to deþ y-brouȝt		4364
Haue mi lyoun	to deþ y-brouȝt		4407
Alle he haþ hem	to deþ y-brouȝt		4784
Hou god him haþ	fram deþ y-brouȝt		5954
Hou Gij hem haþ	fram deþ y-brouȝt		6582
þy ffayre sone	to deþe is brouȝt	RLH	820
Whan þe erle	to deþe was brought	Beves (Cheth.)	2631

A common variant is:

Whi þat icham	in sorwe brouȝt	Guy	286
Now is Gij	in sorwe ybrouȝt		309
þat hire haued	in sorwe brouht	Hav.	336
þat haues him-self	in sorwe brouht		2811
þat he ne weren	to sorwe brouht		57
And haueth me	to sorwe brouht		1372

Many other changes can be rung within the metrical pattern:

þat wimen han	to gronde y-brouȝt	Guy	1564
Now is Beues	to grounde brouȝt	Beves	2783
þan he was	to þe erþe brouht	Hav.	248
Allas! to grounde	icham ybrouȝt	Guy	486
And in þis prisoun	icham y-brouȝt		6198
Whan hii were eft	to gedre brouȝt	Beves	1574

Even more prolific is *gon*. Out of nearly a hundred examples I quote five:

& in to an erber	he is y-go	Guy	565
To an ermite	he is y-go		1647
Toward Seynt Omer	he is y-go		1795
To his felawes	he is y-go		3063
Into þe forest	þai ben y-go		4800

With very little change the infinitive may be substituted:

In to his chaumber	he gan gon	Beves	648
Into the castel	he gan to gon	Guy	678
Into þat schyp	þey gunne gone	RLH	94
To Horn	he gan gon	Horn	1351

Of this formula I count more than thirty examples in the six romances under consideration.

Although the demonstration—that what we have to do with here is the formula in the Milman Parry sense—carries conviction through the quantity of the evidence as well as its quality, there is not space to present the matter in greater detail. One could illustrate the point by citing examples under *in that bataile, that may betide, flesh and bone, that maiden bright, that was so bright, fair and bright, a good brond, sorrow and care, with glad (mild, laughing) cheer, night and day, when it was day, no longer dwell, word and end, well enough, withouten fable, withouten fail, saunfaile, mani and fale, that is so free, at the frome*, and so on through the alphabet.

While the formula finds its greatest use at the end of the line where the composer or reciter must face the requirement of rime, it is also useful in furnishing a ready-made opening for the verse, a general-purpose approach to the main predicate. Here the prepositional phrase with adverbial force has a high utility value. A few examples, taken from the narrator's general stock of such phrases, will illustrate the way they are used:

In þe world	ne worþ man of so gret miȝt	Guy	4263
In the worlde	is not theyr pere	RLH	684
In þe world	was non hym lyche	Ath.	57
In þe world	was non here pere		69

In þe world	is non hys pere		114
In þis world	is better non at nede	Guy	6120
In al þis werd	ne haues he per	Hav.	2241
In al þis world	nis þer man	Beves	1101
In al þis world	nis þer man		1113

Here the idea introduced is generally a comparison—there is no one the hero's equal. On the other hand, the phrase *in his heart* generally announces an emotional state or mood:

In is herte	swiþe bliþe	Beves	1078
In is hertte	him was ful wo		1712
In hys herte	he was ful glad, iwys	(Caius)	p. 133
In his hert	was sorwe anouȝ	Guy	6866
In his hert	him was ful wo		2992
At his hert	him was ful wo		4496
At his hert	gret sorwe sat		6932
And in his herte	made glad chere	Ath.	120
In hert	he is me lef & dere	Guy	2278
In hert	y was glad & liȝt		4640
In herte	he was ful woo	Ath.	81
In herte	he was ful woo		252

In the following examples the various purposes are self evident:

In his hond	a gode swerd he bar	Guy	4131
In his hond	his swerd blodi		5118
In his hond	his swerd kerueing		5238
In his hond	he bar a trounsoun		5272
In his hand	a spere stark	Hav.	380
& in his hond	a swerd wele kerueinde	Guy	5338
& on his hond	his swerd kerueing		4020

To Iesu Crist	he made is pleint	Beves	1576
To Iesu Crist	he bed a bone		1645
To Iesu Crist	he bad a bone		1795
To Iesu Crist	a bed a bone		2629
To Iesu Crist	he gan to calle		2860
To Iesu Cryste	gan he pray	Beves (Chetham)	2508
To Iesu Cryst	he prayde a bone	Ath.	625
To Iesu	he made his praiere	Beves	4415
To god	he made his praiere		2867
To god almyȝty	he bad a bede	Beves (Caius)	p. 134

In many cases the conventional opening of the line is followed by an equally conventional close:

On the morowe,	whan it was daye	RLH	775
On þe morwen,	hwan it was day	Hav.	811
On þe morwen,	hwan it was day		1920
On the morwen,	hwan day was sprungen		1131
On þe morwen,	hwan it was liht		2190
A morwe,	whan hit was dai cler	Beves	755
A morwe,	whan hit was dai cler		3521
A morwe,	so sone so it was day	Guy	4173
Amoreȝe,	þo þe day gan springe	Horn	645
To morwe,	so sone so it is day	Guy	841
Erliche amorwe,	when it was dai	Beves	1973

The descriptions of fighting naturally furnish many examples:

Þourȝ is bodi	went þe dent	Beves	4383
Þurch þe bodi	þe swerd gan gon	Guy	1372
Þurch þe bodi	his swerd glod		1377
Þurch þe bodi	he him smot		1431
Þurch þe bodi	he him smot		4367
Þurch þe bodi	a kniȝt he bar		5073
Þurch þe bodi	he smot him anon		6427
Þurch þe bodi	he him carf atvo		7283
Þurch his bodi	þat swerd ȝede		1439
Þurch his bodi	þe swerd he þriste		1463
Þurch his bodi	þe brond went		5077
Þurch his bodi	þe swerd is gon		6965
Þurch þat bodi	þat swerde bot		7282

and many more examples cited below in connection with the theme of battle.

The consequence of such an action is obvious and is expressed again and again in the same way. By the use of a few well-tried statements, such as *Dead he felled him*, the story-teller is able to give all his attention to the line ending and the necessary rime:

Ded he feld him	doun fot hot	Guy	2946
Ded he feld him	an hast þo		2994
Ded he feld him	adoun þar		5074
Ded he feld him	verrament		5078
Ded he feld him	of his stede		5132
Ded he feld him	in þe feld		5218

Þat ded he feld him	in þe mede		1368
Þat ded he feld him	in þe way		1390
Þat ded he feld him	anon riȝt		2400
& dede he feld him	on þe grounde		2938
Þat ded he feld him	anon riȝt		3516
Þat ded he feld him	on þe ston		5122
Þat ded he feld him	anon riȝt		5220
Þat ded he feld him	on þe sond		5226
Þat ded he feld him	anon		5813

Ded a fel	on þe pauiment	Beves	4384
Ded he fel	wiþ-outen abod	Guy	1378
Ded he fel	wiþ-outen abod		2952
Ded he fel	to grounde þo		7284

Þat ded he fel	doun fot hot	Guy	1880
Þat ded he fel	anon riȝt		2232
Þat ded he fel	wiþouten letting		3522
Þat ded þai fel	mani on		5088

After a dangerous or trying experience which has a happy outcome the hero (or heroine) is naturally "blithe" and thanks God or his worldly benefactor. For such a standard idea a formula is a convenient accessory:

He þankyd god	wiþ deuocyoun	Beves (Caius)	p. 134
And þankyde hem	mani a siþe	Beves	530
And þankede him	ful mani a siþe		906
And þankede god	ful mani a siþe		3472
& þonked god	of þat gras	Guy	1058
& þonked god	ful ȝern, y-wis		4074
And thankyd god	fele sythe	Guy (Cbg.)	772
And thankyd god	fele sythe		5190
And thankyd god	fele sythe		6970
And thankyd god	oftesythe		11588
He thankyd god	in trynyte		11625
They þankyd god	on all manere		11966
And þankede God	ful fele siþe	Hav.	2189
She þanked God	[ful] fele syþe		2843
And þankyd God	off his myȝt	Ath.	590
And þankyd God	off his grace		614
Þey þankyd God	al so blyue	RLH	1249

Travel and movement make up a considerable part of the action of medieval romance, and therefore an adverbial onset in a line is particularly common:

Into þe cite	of Arrascoun him haþ y-diȝt	Guy	1886
Into þe cite	he him diȝt		1992
Into þe cite	þai ben y-gon		1995
Into þe cite	þai gun it bring		4087
Into þe cite	of Lorein . . .		5595
Into þe cite	he is y-go		6343
Into þe cite	of Pauie		6389

A variety of opening phrases can be constructed on this pattern—*And to the cite, To the cite, Out of the cite*, all of which are common—and of course one can enter or leave many other places:

Into þe castel	he gan to gon	Guy	678
Into þe court	he gan to gon		680
In to his chaumber	he gan gon	Beves	648
Out of þat cuntre	he is sone y-go	Guy	6726
And to þe castel	þai wente isame	Beves	3449
Into Almayne	ichil gon	Guy	3245
Into Inglond	he wald wende		1693
Into Cyprys	redy to fare	RLH	626
Into Denemark	for to fare	Hav.	2955
Vnto Londone	wolde he fare	Ath.	89
And to þe gate	Beues ȝode	Beves	2239
Into þe forest	þai ben y-go	Guy	4800
Vnto his in	he is y-go		2845

I have illustrated merely the variety of such phrases without attempting to indicate their frequency. Even so limited a phrase as *into the forest* occurs six times among the quotations I have gathered, and I have not tried to make my collection complete even for the six romances studied. As an exhibit of frequency I may cite:

Forþ þai wente	wiþ þat child	Beves	509
Forþ þai wente	al isame		705
Forþ him wente	sire Beuoun		1133
Forþ him wente	sire Beuoun		1347
Forþ him wente	sire Beuoun		1947
Forþ a wente	be þe strem		1959
Forþ a wente	in is wai		1974
Forþ þai wenten	alle þre		2551
Forþ a wente	to þe castel gate		2977
Forþ þai wente	& drowe saile		3026
Forþ a wente	ase hot		3079
Forþ a wente	also whate		3081
Forþ þai wenten	ase snel		3353
Forþ þai wente	boþe ifere		3635
Forþ he went	into Speyne	Guy	1065
Forþ þai ȝede	wiþ gode welle		5115
Forþ þai wenten	hastily		6568
Forþ þey wente	wiþ glad chere	RLH	621
Forþ þey wenten	wiþouten ensoyne		1475

In view of the frequency of formulas at the beginning as well as at the end of the line it is not surprising to find that the formula sometimes extends to a whole line. Instances occur among the lines quoted in the previous discussion, but a few others may be offered here in evidence. It should be remembered that the substitution of a metrically equivalent word (*so, sone, þat, þo*) is permitted in the formula:

(*a*)	Þat wiþ inne a lite stonde	Beves	733
	And þo in a lite stonde		885
	Sone wiþ inne a lite stounde		1258
	So wiþ inne a lite stounde		1409
	Þo wiþ inne a lite stounde		3489
	So þat in a lite stounde		4393
	Well sone in a lytull stounde	Guy (Cbg.)	1250
	So that in a lytull stownde		3317
	That wythynne a lytull stownde		3363
	And þere wythynne a lytull stownde		10757
	Sone wiþjnne a lytyl stounde	RLH	4565
	So þat wiþjnne a lytyl stounde		6115
	So þat on a litel stund	Hav.	1858
(*b*)	Boþe in yrene and in stel	Beves	2728
	Bothe in yron and in stele	Guy (Cbg.)	960
	Bothe in yron and in stele		1548
	Bothe in yron and in stele		2398
	Bothe in yron and in stele		1644
	Bothe in yron and in stele		2816
	Bothe in yron and in stele		3388
	Bothe in yron and in stele		4976
	Bothe in yron and in stele		6048
	Odur in yron or in stele		8860
	Boþe in yryn and in steel	RLH	2544
	Boþe in yren and in steel		2788
	Boþe in yryn & in steel		5658
	Boþe in yryn & in steel		6048

(*c*) Wiþ his swerd þat wele bote Guy 3562
Wiþ her swerdes þat wil wel bite 2934
Wiþ swerdes þat wil wele bite 5198
Wiþ his fauchon þat wele bote 3556
Wiþ hys fawchoun þat byttyr bot RLH 4874
Wyth a swerde þat wolde wele byte Guy (Cbg.) 1056
Wyth brondys þat full wele cowde byte 7986
Wyth ther swyrdys þat well cowde byte 10960
Wyth my swyrde þat well can byte 11511
Wiþ Morgelay, þat wolde wel bite Beves 3407

It is unnecessary to carry the matter further, although at times a couplet assumes the character of a formula, at least within a single romance. Thus in the Caius manuscript of the *Beves* (pp. 132–134) we find these three almost identical couplets:

A good swerd he gan out breyde,
And on þe dragoun ffaste he leyde. (23–4)

Hys swerd he gan out to breyde,
And on the dragoun ffaste he leyde. (55–6)

Hys swerd he gan out for to breyde,
And on þe dragoun faste he leyde. (95–6)

Again, in the version of *Guy of Warwick* in the Cambridge University Library manuscript (sometimes called *Guy B*) the following couplets occur:

Nowe begynneth newe batayle:
Echon odur faste can assayle. (1027–8)

Then began a grete batayle:
Echon odur faste can assayle. (1825–6)

There beganne a grete batayle:
Euery man odur faste can assayle. (1939–40)

There beganne a stronge batayle:
Eyther other faste dud assayle. (9347–8)

The second line also occurs without the first:

Ayther can odur faste assayle. (5030)

Other instances of such repeated couplets are to be found, though they are naturally not common, and they may be due to the habit of the individual poet.

III

The use of *themes*—recurring features of the description or narrative which are treated according to an established pattern—is one of the characteristics of oral composition, and occurs also in the Middle English romances. Naturally, some of those that are common in the Old French epic have their equivalents in English. I have cited one of Rychner's examples above, the arming of a knight. Such descriptions in the English romances likewise follow a conventional pattern, although the knight is generally described as arming himself. It should be remembered that in the theme verbal identity is not expected. In English the knight girds on a sword, takes a spear in his hand, hangs a shield about his neck or at his side, and rides off:

A gerte him wiþ a gode brond
And tok a spere in is hond,
A scheld a heng vpon is side,
Toward þe wode he gan ride. Beves 759–62

And gerte him wiþ a gode bronde
And tok a gode spere in is honde;
A scheld aboute is nekke he cast
And wente out of þe chaumber in hast. *Ibid.* 1667–70

And gerte him wiþ a gode bronde
And tok a spere in is honde
Out ate gate he gan ride. . . *Ibid.* 2729–31

And gerte him wiþ a gode brond
And tok a spere in is honde,
Aboute his nekke a doble scheld:
He was a kniȝt stout and belde.
On Arondel a lep þat tide,
In to þe strete he gan ride. *Ibid.* 4367–72

Sythen he gurde hym wyth hys bronde:
Hyt was worthe moche londe.
Hys schylde he caste abowte hys halse
And a spere he toke alse.
Hys gode stede he bestrode:
Forthe of the cyte sone he rode. Guy B 3625–30

Seþþe he gert him wiþ a brond
þat was y-made in eluene lond.
His scheld about his nek he tok,
On hors he lepe wiþ-outen stirop,
On hond he nam a spere kerueinde,
Out of þe cite he was rideinde. Guy 3861–6

Another conventional theme involves the question "What is your name?" or "Where do you come from?" and the prompt answer, sometimes with the assurance that the person questioned will not lie or conceal his name. When Horn and his companions arrive on the western strand to which their small boat drifts, the king questions them:

Þo spac þe gode kyng;
He nes neuer nyþyng:
"Sey, child, whet is þy name?
Shal þe tide bote game."
Þe child him onsuerede
So sone he hit yherde:
"Horn ycham yhote,
Ycome out of þis bote
From þe see side:
Kyng, wel þe bitide!" Horn 203–12

In a similar situation Beves is questioned:

"Child," a seide, "whar wer þe bore?
What is þe name? telle me fore!
ȝif ich it wiste, hit were me lef."
"For gode," a seide, "ich hatte Bef,
Iboren ich was in Ingelonde,
At Hamtoun, be þe se stronde;
Me fader was erl þar a while," etc. Beves 539–45

In *Guy of Warwick*, the Cambridge version, we have:

"Syr knyght, telle me beforne:
What ys þy name? where were þou borne?"
"Gye of Warwyk, for sothe, y hyght:
In Ynglonde was y borne aryght." Guy B 629–32

and later he questions Herhaud in similar terms:

"Gode man," he seyde, "what ys þy name?
So god þe schylde fro synne & schame."
"Harrawde," he seyde, "men clepe me
Of Ardurne in that cuntre." *Ibid.* 1339–42

The knight often expresses his willingness to answer:

"Sir kniȝt," he seyd, "y prey þe,
Tel me þi name and whenne tow be."
Sir Gij answerd wel freliche,
"Y schal þe tel ful bleþeliche:
Gij of Warwike men clepeþ me;
Ich was y-born in þat cuntre." Guy 933–8

or says he will not conceal his name:

Quaþ þe soudan, "whennes artow,
Into mi court comen art now,
& misseyst me so schameliche,
& þretest me so dedeliche?"
Gij answerd, "ich-il þe telle:
Mi name for-hele y nille.
Gij of Warwike mi name is;
In þat cuntre y was born y-wis." *Ibid.* 3935–42

In *Richard the Lion-Hearted*, as a preliminary to a proposal of marriage Henry asks the lady her name:

Þenne askyd he þat lady bryȝt:
"What hyȝtest þou, my swete wyȝt?"
"Cassodorien, wiþouten lesyng,"
Þus answeryd sche þe kyng. RLH 171–4

Such a conventional question and answer is of frequent occurrence in the romances and many other examples could be cited. When the poet or minstrel came to a point where an inquiry of this sort was appropriate he knew exactly what to do. In dealing with the theme the story teller follows an established pattern.

The most extended theme is the description of a fight, an encounter between two knights or a battle; the latter is generally a succession of such encounters. The frequency with which this theme occurs varies with the nature of the story. *Beves* and *Guy of Warwick* yield the most examples. In it there are certain recurring elements, not all of which are present in all examples, but all recur with sufficient frequency to indicate that they are part of the general stock upon which the narrator drew. They may be set down in a more or less logical order, as follows:

1. The opponent approaches
2. The hero takes note, or "stood and beheld"
3. He leaps into the saddle
4. He spurs his horse
5. He shouts an angry defiance or orders his opponent to yield
6. He rides against the foe
7. The knights meet head-on; or
8. They strike each other
9. Fire flies
10. One strikes the other's shield
11. Part of the shield (helmet, hauberk) flies off
12. The hand goes too
13. He strikes the helmet
14. He strikes the head
15. He strikes the head off, or it flies off
16. He strikes the neck, shoulder, side, etc.
17. The spear breaks
18. He unhorses his opponent
19. He fells horse and man
20. Both are unhorsed
21. He draws his sword
22. The sword bites well
23. He strikes with sword
24. They lay on eagerly, fight "as they were wood," spare not
25. He cleaves opponent to the saddle, chin, etc.

26. He cuts the body, etc. in two
27. He kills opponent or gives a deadly wound
28. He pierces him through the body, heart, etc.
29. He fells him to the ground, or
30. The enemy falls to the ground (dead)
31. Armor availed not
32. He taunts the vanquished foe
33. "With that came . . ."
34. He smote him
35. The victor returns home, or departs, with or without prisoners

In addition to these rather specific details of the action there are certain subsidiary features or types of general comment:

36. The knight is stung by injury or anger to retaliate
37. He is "full woe," "in heart wroth," etc.
38. The knight is named or characterized parenthetically
39. Size of host stated
40. How long they fought
41. Why tell a long story?
42. Summing things up
43. None could stand against him (them)
44. The hero slays many
45. Prisoners are taken
46. The enemy flees, or
47. None escaped alive
48. Never before was such a fight
49. The victors give thanks

The evidence on which the above enumeration is based appears in the appendix. The reader who is anxious to get along with the argument may disregard the quotations there assembled, or sample them according to his taste. For these are the keys which the narrator can press and receive from his mental storehouse an idea suitable at the moment. Since it is an idea or motif that he is summoning, he is not bound to express it always in identical words but is free to vary it to suit the context, especially the rime. It is interesting to note, nevertheless, how frequently the same words are used. It is not necessary to point out that not all the forty-nine elements are used in any one occurrence of the battle theme, nor are those selected employed in a fixed sequence. Rather the story teller may select freely from the stock and assemble the units in any way he chooses. We may illustrate the procedure by one battle scene from *Guy of Warwick*. A number of others would serve equally well, but this has the virtue of being relatively brief. In the fighting Tirri is being hard pressed and Guy and Herhaud go to his relief. I indicate within brackets the numbers of the elements in the previous list as they occur in the passage. The passage represents lines 5117–58.

Wiþ þat com forþ sir Gij, [33]
In his hond his swerd blodi. [38]
Wel heteliche he smot a kniȝt, [34]
His bodi he clef adoun riȝt. [26]
Anoþer kniȝt he smot anon [34]
Þat ded he feld him on þe ston. [29]
Sir Gij him smot to Gayer, [34]
Þat was þe doukes nevou Loyer: [38]
Of his hors he haþ him feld [18]
Þurch Tirries help in þe feld.
For he smot his felawe, [36]
In þe sond he haþ him slawe. [27]
Herhaud smot anoþer forþ, [34]
His armes was him nouȝt worþ: [31]
Þurch his bodi þe swerd ȝede; [28]
Ded he feld him of his stede. [29]
Anoþer he smot him as gode kniȝt, [34]
Of his stede he feld him doun riȝt. [18]
Now þai ginne togider smite: [8]
Non no spared oþer bot lite. [24]
Togider þai smite hard wiþ alle [8, 24]
Mani on þer was ded & doun falle. [30]
Who þat seye þan þerl Tirri
Wiþ his felawe sir Gi,
& Herhaud of Arderne þe gode,
Þat wele to smite was in his mode,
So mani þai nomen & feld þat day
Is non þe best chese may
Of þre kniȝtes so wele doinde,
Oȝain þe Loreyns fiȝting. [42]
Gij to þe steward haþ y-smite, [34]
Of his hors he feld him wiþ hete: [18]
Þurch strengþe of fiȝt he him wan.
Þe oþer oway fleye ich man: [46]
Gij & Tirri hem folweþ strongliche
& her feren, hardiliche.
Alle þai ben ded oþer ouer-come: [44]
Þer bileued non vnnome. [45]
Gij & Tirri oȝain ben y-gon: [35]
Wiþ hem þai ladde her prisouns ichon. [45]
Whar-to schuld y þer-of lye? [41]
Þat day þai hadde þe maistrie. [42]

IV

The formula and the theme are the features of improvisation upon which most stress is laid in discussions of the Greek, Jugoslav, and Old French epic. There is, however, in the English romances another which seems to be somewhere between these two and to be of equal importance.

It is of very frequent occurrence and may be called the *predictable complement*. Certain statements seem to call up automatically in the mind of the poet or reciter a conventional way of completing the thought. It is as though he were subject to a kind of conditioned reflex. Generally the statement and its predictable complement form a couplet, and this feature of the composition is the result of the fact that the couplet is the basic unit of most English romances, even the stanzaic romances. Here couplets are linked by tail lines, whether the stanza is the romance six or the twelve-line tail-rime stanza, a combination of two romance sixes. Considerations of rime are doubtless responsible for many a predictable complement, but are not the complete explanation. Such complements occur not only where rime words are not easy to find but in cases where there is a considerable range of choice. We may illustrate this, as well as the feature itself, by considering the word *knight*, which naturally occurs with great frequency in the romances.

The word *knight* is not a difficult rime word. It rimes with *dight*, *fight*, *hight*, *light*, *might*, *night*, *pight*, *right*, *sight*, *tight* (verb), and *wight* (noun and adjective), and all these rimes are actually found in the romances I am considering. But three of them occur with such frequency as to suggest that the narrator tends to fall into a few conventional thought patterns. For example, upon mention of a knight (or knights) it is very common to add that he was good in a fight:

On eueryche lepte an Englysshe knyght,
Stowte in armes, and stronge in fighte. RLH 2275–6

And bad hym take wiþ hym knyȝtes,
Stoute in armes, stronge in ffyȝtes. RLH 5177–8

He made here warde of noble knyȝtes,
Stoute in armes, stronge in ffyȝtes. RLH 5945

Þorwȝ help off hys gode knyȝtys,
Stoute in armes and stronge in fyȝtes. RLH 6503–4

So þat ichaue fourti kniȝtes,
Stout in armes & strong in fiȝtes. Beves 3605–6

It happens that in the five illustrations the second line of the couplet is practically the same throughout. But the idea can be expressed with variations in wording:

Afftyr comen barouns and knyȝtes,
An hundryd þousand stronge in fyȝtes. RLH 2979–80

Þai weren boþe strong kniȝtes,
Bold and hardi in ich fiȝtes. Guy 779–80

Of euerich londe þider com kniȝtes,
Þat strong ben & bold in fiȝtes. Guy 813–4

Þan com Torald, a gode kniȝt,
Swiþe gode & hardi in fiȝt. Guy 1379–80

& wiþ him oþer fifti kniȝt,
In feld þe best þat miȝt fiȝt. Guy 1919–20

Wiþ fif hundred armed kniȝtes,
Hardi & wele doand in fiȝtes. Guy 2323–4

Sir Giȝ toke an hundred of his kniȝtes,
Strongest and best in fiȝtes. Guy 2861–2

Wiþ him he haþ an hundred kniȝtes
Of Almayne, þe best in fiȝtes. Guy 3127–8

He was coraious & gode kniȝt,
& michel adouted in euerich fiȝt. Guy 3465–6

& wiþ hem fif hundred kniȝtes,
Orped men & gode in fiȝtes. Guy 6061–2

Many other changes are rung on the rime *knight-fight*, which would only serve to demonstrate the ingenuity of the poet or the variety of stock phrases available to carry out the same sequence of thought.

The adverbial phrase *anon riht* is a formula that serves as an easy rime in many couplets. In *Guy of Warwick* it is a favorite in couplets where in one line a knight is struck and in the next line falls dead:

Wel sternliche he smitt a kniȝt,
Þat ded he fel anon riȝt. Guy 2231–2

Wel hardeliche he smot a kniȝt,
Þat ded he feld him anon riȝt. Guy 2399–2400

He forþ ȝede & smot a kniȝt,
Þat ded he feld him anon riȝt. Guy 3515–6

His feren he rescoud as a gode kniȝt,
Mani on he feld ded anon riȝt. Guy 5107–8

Anon he smot anoþer kniȝt
Þat ded he feld him anon riȝt. Guy 5219–20

The same sequence of ideas is apparent in

Wel heteliche he smot a kniȝt,
His bodi he clef adoun riȝt. Guy 5119–20

Anoþer he smot him as gode kniȝt,
Of his stede he feld him doun riȝt. Guy 5133–4

As we have just seen, the hero sometimes takes a hundred or five hundred knights, hardy or strong in fights. But the second line may instead indicate that he sets out at once against the enemy, again with the rime *anon riht:*

He tok wiþ him sex kniȝtes
And wente forþ anon riȝtes. Beves 4291–2

Nim we now an hundred kniȝtes,
& go asayl hem anon riȝtes. Guy 2337–8

For ther were a thousand knyghtes
With theim to mete anon Rightes. Guy 2109–10 (Caius)

& wiþ him went alle þe kniȝtes,
Acord to make anon riȝtes. Guy 6621–2

and with a slight variation of the adverbial phrase

Mani was þe gentil kniȝt
Þat wiþ hem went þo riȝt. Guy 6695–6

But the rime *kniȝt-anon riȝt* lends itself to a variety of situations:

Whi slouȝ ȝe þat ich kniȝt?
Alle ȝe schul die anon riȝt. Guy 4705–6

Þe bodi þai toke of þat kniȝt,
Opon a pal leyd it anon riȝt. Guy 4845–6

Wiþ fif hundred of gode kniȝtes:
An acumbraunce ous come anon riȝtes. Guy 5165–6

When Gij him seye he knewe þe kniȝt,
He kist him þer anon riȝt. Guy 6041–2

With such a universally applicable adverbial phrase it is to be expected that inversion would be easy and therefore common:

King Yuor grauntede anon riȝtes;
He let him chese fourti kniȝtes. Beves 3611–2

Josiane þo anon riȝtes
Clepede to hire twei kniȝtes. Beves 673–4

King Ermin þo anon riȝte
Dobbede Beues vn-to kniȝte. Beves 969–70

And sai þemperur anon riȝt,
Þat i nam no Frensche kniȝt. Beves 3065–6

And armede þer anon riȝtes
Boþe he and is sex kniȝtes. Beves 4365–6

And kiste hire anon riȝt
And sente after baroun & kniȝt. Beves 3171–2

Wiþ þat stirt forþ anon riȝt
Otus cossyn, an vnwrast kniȝt. Guy 5733–4

In all these cases it can hardly be doubted that the pattern in the narrator's mind is the rime *anon riȝt-kniȝt*, rather than any conventional sequence of ideas, and perhaps these cases should not be cited here. They represent, however, an associated phenomenon, and may have come about through the rime pattern in such cases of predictable complement as I have just cited.

Since strength and courage are such indispensable attributes of the knight, it was inevitable that the association of *wiȝt* and *kniȝt* should appeal to some poets. It was a favorite of the *Havelok* author:

Thi-self shal dubben him to kniht,
For-þi he is [man] so wiht. Hav. 2042–3

And so shaltu ben maked kniht
With blisse, for þou art so wiht. Hav. 2186–7

In þis middelerd nis no kniht
Half so strong, ne half so wiht. Hav. 2244–5

and with inversion of the rimes

Bernard was trewe, and swiþe wiht,
In al þe borw ne was no kniht. . . . Hav. 1756–7

For y se þu art so wiht,
And of þi bodi so god kniht. Hav. 2720–1

The prowess of a knight may also be symbolized by the word *might:*

Now is sir Gij dobbed to kniȝt;
Feir he was and michel of miȝt. Guy 723–4

He rode to him as a gode kniȝt,
He semed a man of miche miȝt. Guy 873–4

Anon þai nomen an hundred kniȝtes,
Hardi & of most miȝtes. Guy 2347–8

The following couplet shows inversion of the rimes, but the lines are the same syntactically in either order:

So strong he is & of so gret miȝt,
In world y wene no better kniȝt. Guy 2907–8

But *might* may also be a verb:

Of al þe londe þe stringeste kniȝte,
Þat hii owhar finde miȝte. Beves 3309–10

Neuer no was an-lepy kniȝt
Þat so mani stond miȝt. Guy 3991–2

Finally, the association of *knight, might,* and *fight* is neatly represented in the couplet

A kniȝt he was of gret miȝt,
Swiþe gode & hardi in fiȝt. Guy 1419–20

There is not space to deal with all the conventional sequences of ideas in such couplets, but it may be noted in passing that in *Guy of*

Warwick the author seems to like to open the second line with a phrase indicating nationality or geographical origin:

Gauter come prikeing anon riȝt,
Of Almayne a wel gode kniȝt. Guy 2953–4

& seþþe he slouȝ a Freyns kniȝt,
In Bleyues he was born ariȝt. Guy 2981–2

Wiþ þat come anoþer kniȝt
Of Fraunce y-bore, Amori he hiȝt. Guy 5327–8

Wiþ þat com forþ þe steward liȝt,
A Brabasone he was, a wel god kniȝt. Guy 6899–6900

One of the predictable complements of greatest frequency is associated with the description of combats whether between individual adversaries or collective forces. A blow is delivered with sword or lance and the victim is felled by it in conventional phrasing. Instances occur among the illustrations already cited. But while I have given a number of examples involving the rime *kniȝt-(anon) riȝt*, the complement is not limited to this rime. Indeed, nowhere do we find better evidence that we are dealing not with a particular rime pair but with an almost inevitable sequence of ideas than in the great variety of rimes involved:

Wiþ grete strengþe sir Gij him smot
Þat he feld him anon fot hot. Guy 3509–10

To þe douk Segyn he smot,
& of his hors feld him fot hot. Guy 2383–4

Gij þe steward so hard smot,
Of his stede he feld him fot hot. Guy 1957–8

Gij to þe steward haþ y-smite,
Of his hors he feld him wiþ hete. Guy 5147–8

Sir Gij him smot to Gaier,
& feld him doun of his destrer. Guy 2355–6

I smot hym þat alle folk it seyȝ,
Doun off hys hors almost he fleyȝ. RLH 545–6

Ac a failede of his diuis
And in þe heued smot Trenchefis,
Þat ded to grounde fel þe stede. Beves 1887–9

With a dente amyd the shelde;
His horse he bare downe in the felde. RLH 291–2

& mani anoþer he haþ aqueld,
& adoun feld in þe feld. Guy 2995–6

Ermine he smot on þurch þe scheld;
Almost he feld him in þe feld. Guy 3499–3500

Þe bodi atvo he haþ to-deled,
Þat he fel doun in þe feld. Guy 2957–8

(where the consequence is not surprising)

Herhaud him haþ þer afeld,
Þat dede he lay in þe feld. Guy 5229–30

Tirri anoþer smite bigan,
Þat ded he feld boþe hors & man. Guy 5479–80

& smot him so vpon þe croun,
Þat man and hors de cleuede doun. Beues 4513–4

And smot him so up-on þe crune,
Þat Godrich fel to þe erþe adune. Hav. 2734–5

Vpon is helm in þat stounde,
Þat a felde him flat to grounde. Beves 1039–40

Vpon þe helm in þat stounde,
Þat man and hors fel to grounde. Beves 3433–4

To a Lombard smot sir Gij,
& feld him & his fere him by. Guy 5477–8

Þurch þe bodi he him carf atvo:
Ded he fel to grounde þo. Guy 7283–4

Wiþ his fest he him smot so,
Þat to grounde he dede him go. Guy 5789–90

And on þe geauntes brest a wonde,
Þat neȝ a felde him to þe grounde. Beves 1909–10

And smot ato his nekke bon:
Þe geaunt fel to grounde anon. Beves 1919–20

In twoo he brak hys cheke-bon;
He fel doun ded as ony ston. RLH 797–8

Rychard hym smot, forsoþe to say,
Euene in twoo hys cheke-bon.
He ffyl doun ded as ony ston. RLH 866–8

Þourȝ is bodi wente þe dent,
Ded a fel on þe pauiment. Beves 4383–4

And [the dragon] smot his hors wiþ þe taile
Riȝt amideward þe hed,
Þat he fel to grounde ded. Beves 2780–2

Þat his heued he him to-clef:
Al to ded to ground he dref. Guy 6905–6

Þemperur he smot wiþ is spere,
Out of is sadel he gan him bere
And þrew him to grounde. Beves 232–4

Gij him smot oȝain, no miȝt he as nauȝt,
Þat he haþ þe grounde y-rauȝt. Guy 6977–8

Þer-wiþ he smot Ebban þe king,
Þat ded he fel wiþouten letting. Guy 3521–2

A little more novelty is introduced in *King Horn* when we are told

> Þe sarazyn he hitte so
> Þat is hed fel to ys to. Horn 605–6

Although this list of examples is a long one, I have omitted more than I have included. Moreover, it could be greatly extended by in-

stances from other romances. They would only prolong what has already reached wearisome length, the proof that given certain statements of frequent occurrence in the first half of the couplet, one can predict with reasonable assurance the idea of the second half. The narrator can complete the thought—one might almost say, without thinking—by drawing upon a stock of ready-made phrases which vary only enough to provide a choice of rimes.

V

In the foregoing discussion I have examined a small group of English romances from the point of view of the theory of oral composition as that theory has been developed in connection with the Homeric poems and recently applied to some early chansons de geste. That theory rests first of all upon the presence of stock expressions which fit a particular metrical need—the *formulas* in Milman Parry's terminology—and, secondly, the presence of certain descriptive or narrative elements, called *themes*, which in their repeated occurrences follow conventional patterns. I think it is beyond question that both of these features are present in the Middle Englisl. romances examined. Are we then to conclude that the same inference is to be drawn here as has been drawn by those who have developed the theory in connection with the Homeric poems and advanced it in connection with certain chansons de geste? Are we to believe that those English romances which meet the test of formulaic and thematic elements were composed in the act of oral delivery as are the heroic poems of Jugoslavia? In its full form I do not see how such a conclusion can be accepted.

I hold no brief for or against the theory of improvisation in the Homeric poems. We can never hope to know what lies back of these poems, what Homer or numerous Homers had to work with. But science proceeds by the constant testing of hypotheses, and the theory that certain features of the Homeric poems can be explained by the observed practices of Jugoslav singers of heroic poems at the present time may be rejected by the skeptical but cannot be disproved. Likewise when we turn to the origin of the chansons de geste we are in almost equal darkness. Whether we believe in the theory of cantilenae or epic lays or in the partnership of jongleur and cleric we do not know what lies behind the *Chanson de Guillaume* or *Raoul de Cambrai*. While the supreme art of the *Roland* (as of the *Iliad*) will prevent many from accepting the theory of oral composition, the theory cannot be disproved in the absence of any knowledge of the literary sources of the heroic poetry of medieval France. But for the Middle English romances the situation is quite different. Many of the Middle English romances are translations or else adaptations of French poems. In cases where the French original can be identified, at least as to its approximate form, the similarities are often such as to leave little doubt that the English poet was following his source with reasonable fidelity, such fidelity as to suggest that the source lay open in a manuscript before him.

The version of *Beves of Hampton* in the Auchinleck manuscript and its four congeners is an interesting text to study from the point of view of oral improvisation. It contains a considerable number of verbal clichés. If we could assume that the minstrel in reciting the story was also partially composing it as he went along, many of these clichés would qualify as formulaic elements. But there can be little doubt that we have to do here with a translation, an English rendering of a French text. There are first of all references to a source—"seiþ þe bok" (844, 2468), "so hit is fonde in frensche tale" (888), "þe romounce telleþ" (1537), "So hit is in Frensch y-founde" (1782). There are still other allusions to the French source in the Cambridge manuscript. While such references could be conventional, I have found that in general when they occur in an English romance they are likely to be confirmed by other evidence. And that happens to be the case in the present instance. There are four versions of the *Beves* romance in Old French (disregarding the version in prose), differing so from one another that Stimming was compelled to publish them separately. One is Anglo-Norman, the other three Continental. (The A-N is in vol. 7 of the *Bibliotheca Normannica*; the others are in the series published by the Dresden Gesellschaft für romanische Literatur, vols. 25, 30, 34, 41–42.) A very little examination of these four texts suffices to establish the fact that the English version we are considering is based on the Anglo-Norman, which happens to be the earliest of the French versions. While the English is a free paraphrase, it follows the French text so closely as to leave no doubt of the relationship, and there are verbal correspondences which are conclusive. The relationship may be indicated by a few stanzas at the beginning.

A-N *Boeve de Haumtone*
(ed. Stimming, Bibliotheca Normannica, VII)

Seingnurs barons, ore entendez a mei,
si ws dirrai gestes, que jeo diverses sai,
de Boefs de Haumtone, li chevaler curtays,
ke par coup de espeie conquist tant bons reys.
Si vus volez oyer, jeo vus en dirrai;
unkes ne oistes meyllur, si com jeo crai.

Seingnurs, si de lui oyer desirez,
Jeo vus en dirrai, kar jeo sai asez;
primes vus en dirrai de soun parentez.
A Haumtone fu li quens plein de bontez,
il out a noun Guioun, chevaler fu prisez;
meilour de lui ne fust en son tens trovez.

Seignurs, iceo quens Guioun dount vus chaunt
estoit bon chavaler, pruz e combataunt;
mes de une chose lui alout home blamaunt,
k'ainz ne vout femme prendre en tot son vivaunt,
dunt pus se repenti par le men ascient.

Mes quant il fu veuz home e out long tens vescu,
donk prist il femme que de haute gent fu,
file au roi de Escoce cele dame fu.
Guioun la prist a femme, lui chevaler membru.
Puis avint cel jour que mult iré en fu,
ke il perdi le chef par desus le bu.

Beves of Hampton (Auch. MS)
(ed. Kölbing, EETSES, XLVI)

Lordinges, herkneþ to me tale!
Is merier þan þe niȝtingale,
þat y schel singe;
Of a kniȝt ich wile ȝow roune,
Beues a hiȝte of Hamtoune,
Wiþ outen lesing.

Ich wile ȝow tellen al to gadre
Of þat kniȝt and of is fadre,
Sire Gii:
Of Hamtoun he was sire
And of al þat ilche schire,
To wardi.

Lordinges, þis, of whan y telle,
Neuer man of flesch ne felle
Nas so strong,
And so he was in ech striue,
And euer he leuede wiþ outen wiue,
Al to late and long.

Whan he was fallen in to elde,
þat he ne miȝte him self welde,
He wolde a wif take;
Sone þar after, ich vnderstonde,
Him hadde be leuer þan al þis londe,
Hadde he hire forsake.

But there is still more interesting evidence. The English poem undergoes a change of metrical form at line 475 of the Auchinleck manuscript. At this point the meter changes from a six-line stanza ($aa^4b^3cc^4b^3$, occasionally $aa^4b^3aa^4b^3$) to four-stress couplets. Kölbing in editing the romance noted the change but confessed that he could not account for it, saying that there was nothing "to correspond with this change in the original French versions" (EETSES, LXV, p. xi). In this he appears to have been mistaken. The Anglo-Norman version does undergo a change in the versification. It is written in laisses like other chansons de geste. Of the 3,850 lines of which it consists the first 2,338 are in laisses in which the final syllables rime, but in the remaining 1,512 lines the laisses are bound together by assonance. But there is a more significant change. As Sarrazin, the editor, notes, the length of the laisses undergoes a very noticeable change after laisse 66 (which ends at line 415). Up to this point they have all been short, generally six lines, with an occasional divergence to five or seven. The effect is that of a poem written in six-line monorimed stanzas. From this point on the laisses are generally much longer, extending in an extreme instance to 187 lines (with assonance). Now the change from monorimed stanzas averaging six lines occurs at very nearly the same point in the story as that at which the English poem changes from six-line stanzas to couplets. While the English adapter's stanzas do not always stand in a one to one relation with the stanzas of the French, since the English poet occasionally omits a stanza, sometimes expands a stanza to two, or even interpolates a stanza of his own (which may, however, have been represented in the particular manuscript he was using), one cannot doubt that he was working with a written text before him. Oral composition would be conceivable only on the assumption that he was truly bi-lingual and could have sung or recited the story in French as well as English. This is conceivable but the less likely of the two possibilities.

It is equally possible to show the close dependence of *Guy of Warwick* upon the French, although the problem is more complicated. In one respect it is simpler. So far as we know

there is only one version in French verse, and this has fortunately now been edited from all the manuscripts.[18] On the other hand there appear to be four versions, in whole or part, in Middle English. For my present purpose one of these will suffice; it is in couplets and is in many ways the best. We may call it the Caius-Auchinleck version.[19] It is so close to the French text that the two often correspond line for line. Although the relationship is beyond question, I quote for convenience a few lines. It is the passage in which Guy, having met Felice's requirement that he prove himself as a knight, returns to ask the fulfillment of her promise:

Guy of Warwick

French text, ll. 1039–54

Gui a Felice s'en est alé;
Mult ducement l'ad salué:
"Venuz sui, ma bele amie,
Par vus ai certes la vie;
Se ne fuissez, jo fuisse morz,
Destruiz e malbailliz del cors.
Les armes prendre me feistes,
Vostre pleisir puis me deistes:
Quant jo les armes pris avreie,
E la mer passé serreie,
E loinz en estrange regné
D'armes fuisse bien preisé,
Granté me serreit l'amur de vus,
Dunt ai esté tant desirus;
Ore sui venuz pur oir,
Bele amie, vostre plaisir."

Caius MS, ll. 1115–30

On a daye he is to Felice goo,
And full louyngly he seith hir too:
"I am comme as thou may see.
My lif y haue, lemman, thurgh thee:
Ne were thou, lemman, dede y were,
My body destroied and leide on bere.
Armes y toke for loue of the,
Thoo thy wille thou tolde me,
That, whan y had armes take,
Thou woldest not than me forsake.
And thou hast herde me preised bee
In many astraunge contree:
Thy loue shuld not me bee werned,
For y haue it, me thinketh, ayerned.
Sweting, nowe y am comme to the,
Thy wille therof thou telle me."

Auchinleck MS, ll. 1115–30

To Felice þan sir Gij is go;
Sweteliche he seyd hir to:
"Leman," he seyd, "wele þou be,
Mi liif ichaue for loue of þe;
Ded ich were ȝif þou nere,
Mi bodi destrud and leyd on bere.
When þou þi wille hadde seyd to me,
Armes y fenge for loue of [þe]
& when ich hadde armes take,
þou seyd þou noldest me forsake,

þou noldest þi loue werne to me;
& nov ich am her comen to þe:
Dere leman, y prey þe
þi wille þatow tel to me."

With respect to the other romances under consideration we are not so fortunate. No source has been found for *Richard the Lion-Hearted*, although there are many references in it to "the book," which the editor, Karl Brunner, assumes

[18] *Gui de Warewik, roman du xiii^e siècle*, ed. Alfred Ewert, 2 v., Paris, 1933; *Classiques français du moyen âge* **74–75**. The quotations from the English *Guy of Warwick* are from the Auchinleck MS unless otherwise noted. In Zupitza's edition (EETSES, XLII, XLIX, LIX) the line numbers 1905–2006 are repeated. Numbers of the second sequence are here followed by an asterisk.

[19] The Caius College MS is the most complete. It is paralleled by the Auchinleck MS as far as line 7306, and there is a fragment of one leaf in Sloane MS 1044. It is edited by Zupitza in EETSES, XLII, XLIX, LIX. The second text is in twelve-line tail-rime stanzas, also preserved in the Auchinleck MS, and begins where the couplet version in this manuscript stops. A third version, in couplets, is only known from three brief fragments. The fourth, likewise in couplets, is in a manuscript in the University Library, Cambridge. This is edited by Zupitza in EETSES, XXV–XXVI. However, two long passages totaling more than 2,000 lines in this otherwise independent version parallel the text in the Caius MS.

to be a lost romance in French. The *Havelok* story is known in two Anglo-Norman forms, but neither is the source of the English romance. Nor is the Anglo-French *Horn et Rimenhild* the source of *King Horn*. No other version of *Athelston* is known.

Thus of the six romances under consideration in this paper we are able to study only two in relation to their source, but these two—*Beves of Hampton* and *Guy of Warwick*—furnish convincing proof that they were not originally the products of oral composition.

VI

Are we then to dismiss the whole question of improvisation from our minds and to regard the presence of large numbers of formulas and themes in English romances as proof only of the ineptness of the poets who composed them? If this were the only conclusion to which the present study leads, I am afraid I should have done no more than convince you that Middle English poets had no monopoly on the insignificant. But it has not been my purpose to lead you by a circuitous route to so unrewarding a destination.

In seeking the answer to our question we must approach the problem from the point of view of medieval conditions and not those to which we are accustomed, and you must permit me a few platitudes. A basic feature of those conditions was a listening audience rather than one which took in the written word through the eye. While literacy outside the clerical class was doubtless spreading at the end of the Middle Ages, it was still quite limited, and anything that could be called the reading habit was nonexistent. Even if the ability to read were more general than it apparently was, the scarcity and expensiveness of books would have put a severe limitation upon its exercise. In other words, we must think of the purveying of literary entertainment as mainly oral—the singing, recitation, or reading aloud of a minstrel or some one performing his function. This is not the place to present the evidence for these three types of performance, but unless the words *sing*, *tell*, and *read* have no meaning in Middle English I believe we must conclude that all three methods were on occasion employed. But once again we must be careful not to attribute modern meanings to medieval words. Anyone who has read the late Professor Chaytor's *From Script to Print* will be slow to think of the reading process in the Middle Ages as involving the easy and rapid transference of written symbol to word and idea that characterizes the modern educated person. In many cases a book would only have assisted the minstrel's memory.

Now if we start with the fact that the medieval minstrel singing or reciting the story of *Guy of Warwick* or *Beves of Hampton* must have been doing very much what the singer of a similar Jugoslav poem does, we may well inquire whether his performance involved a similar combination of memory and improvisation, including the use of stock formulas and themes. And if there is an *a priori* probability that this is so, is there any specific evidence that can be used to test this probability?

It is not uncommon in discussions of the Middle English romances to talk about *Guy of Warwick*, *Beves of Hampton*, or *Richard the Lion-Hearted* as we talk about other long poems like *The Owl and the Nightingale* or, let us say, the A-text of *Piers Plowman*, and thus give the impression that we have to do with a single literary composition which has undergone merely such alteration and corruption as results from the conscious tampering or unconscious lapses of medieval scribes—modernization of language or idea, modification of dialect, or mistakes of transcription, and the like. But with respect to romances I believe this is often not the case.

And here it is necessary to pause long enough to point out that the Matter of England romances, which we are here alone considering, fall into two quite distinct groups. There is the relatively short and fairly unified story like *Havelok* and *King Horn* and *Athelston*, and there is the long narrative made up of an endless succession of individual adventures in which the principal character or one of the main characters is involved. *Beves of Hampton*, *Guy of Warwick*, and *Richard the Lion-Hearted* are examples of the latter type. A characteristic of these romances of loosely strung adventures is their repetitiousness. The same situation—for example, the combat of the hero with an opponent, the free-for-all fighting of a battle in which the hero engages all comers—occurs over and over and lends itself to conventional narrative patterns. For some reason the shorter and more unified type has survived in fewer manuscripts and fewer "versions." *Havelok* and *Athelston* are known from only one manuscript each, and *King Horn* survives in only three. On the other hand there are seven manuscripts and one early

printed text of *Beves*, six texts (counting fragments) of *Guy of Warwick*, and eight of *Richard the Lion-Hearted*. Moreover, the variations between the manuscripts containing these longer narratives are often quite striking, so great that it is generally impossible to establish a single critical text with the variants adequately reported in footnotes. Indeed, the attempt to do so has at times obscured the extent to which the different manuscripts differ from one another except where these differences are so extensive as to force the editor to print long passages independently in the notes. If it were not for certain bookish productions, one might venture the observation that the number of what are really separate versions of a story is in direct proportion to the length of the story.

As I see it, the answer to our question depends upon a study of all the versions of a given text, considering both the lines that are identical (which show that we are dealing with a single poetic composition) and the variations from manuscript to manuscript of that composition (which we must try to account for). If we find that many of these variations cannot be explained by scribal error, alterations such as scribes are known to introduce, or in short by any of the principles of text transmission, we are entitled to ask whether some other explanation accounts for the facts in a satisfactory way. And we must be prepared to do this without thinking in terms of a later time or a different set of conditions. Such a study must be based on romances which have come down to us in more than one manuscript—the more manuscripts the better. And this requirement is best fulfilled among the Matter of England romances by *Beves of Hampton*, *Guy of Warwick*, and *Richard the Lion-Hearted*. Such a study could embrace the whole of a romance, but fortunately the results can be presented rather briefly. If the evidence which I consider significant really exists, it can be seen as well in a hundred lines as in a thousand, often indeed in a relatively brief passage. I shall have to confine myself here to one exhibit, or at most two, from each of the romances just mentioned.

We may begin with *Beves of Hampton*. The romance is an excellent one for such an investigation since all the texts are clearly descended from a single original. I select a short episode which occurs in all seven texts. Beves and Josian are riding through the woods in France, accompanied by Beves's cousin Terri, when they have to stop in order that Josian may give birth to a child. They build her a lodge and then discreetly go some distance away "For hii ne miȝte hire paines here." She is delivered of twins. Unluckily Beves has gone too far and Josian is almost immediately kidnapped and carried off. When Beves and Terri return they find only the two children. Wrapping them in ermine cloaks they ride until they meet with a forester, to whom they commit one of the infants. The passage as it appears in the Auchinleck manuscript is as follows (ll. 3723–40):

> Þar nolde hii no long abide,
> Þei lope to horse & gonne ride;
> In þe wode a forster þai mette
> And swiþe faire þai him grette:
> "God þe blesse, sire!" Beues sede,
> "Siȝe þe eni leuedi her forþ lede
> Owhar be þis ilche way?"
> "Sire, for gode," a seide, "nay!"
> "What dones man ertow, bacheler?"
> "Sire," a seide, "a forster!"
> "Forster, so Crist þe be milde,
> Wiltow lete cristen þis heþen childe?
> Riȝt, lo, now hit was ibore
> And ȝong hit haþ is moder for-lore:
> Wilt þow kep it for [to] min," a sede.
> "And i schel quite wel þe mede?"
> Þe forster him grauntede þer,
> To kepe hit al þe seuen ȝer.

It is not possible to present the seven texts of this passage in parallel columns, and so I resort to the device of printing each line separately, followed by the corresponding line in the other manuscripts. The order of the manuscripts is always the same: Auchinleck (A), Cambridge (Cbg), Chetham Library, Manchester (M), Caius College (Ca), Egerton (E), Naples (N), and the Pynson print (P). Since E and N are practically identical, wherever possible I print only E with the variants of N in parentheses:[20]

3723. A. Þar nolde hii no long abide,
Cbg. Wolde þey no lenger byde,
M. No lengere they wold abide,
Ca. No lengere wolde þey abide,
E(N). And no lengere nolde he (þey) abide,
P. No lenger there wolde he abyde,

[20] I have had to depend on Kölbing's notes for the variants and since he records only changes of wording, the spelling is not necessarily exact.

3724. A. þei lope to horse & gonne ride;
Cbg. But toke þer horsys & forþe can ride;
M. But lope on horse & forþe can ride;
Ca. But lope on horse & gonne to ride;
E(N). But lept on horse for (& gonne) to ride;
P. But toke his chyldren and forthe gan ryde.

3725. A. In þe wode a forster þai mette
Cbg. In þe fforeste a man þai mette
M. In þe wode a forster þai mette
Ca. In þe wode forster þai mette
E(N). *like* A
P. A foster in the wode he met.

3726. A. And swiþe faire þai him grette:
Cbg. And feyre sir Befus him grette:
M. And Beues hym ffull ffeyre grette:
Ca. *like* A
E. And sir Beues feire him grette:
N. Swiþe sir Beues wel feire him grette:
P. And syr Beuys fayre hym gret:

3727. A. "God þe blesse, sire!" Beues sede,
Cbg. "Crystys blessyng mote þe betyde
M. "Cryste mvst the blesse," he said
Ca. And sayden to hym: "So god þe spede
E(N). "Crist þe blesse," he seide þan
P. "Felawe," sayde he to hym than,

3728. A. "Siȝe þe eni leuedi her forþ lede
Cbg. "Siȝe þe eni wumman þys way ryde
M. "Siȝe þe eni man a wumman lede
Ca. "Siȝe þe owher a wumman lede
E(N). "Siȝe þe eni wumman her forþ gon
P. "Sawest thou ought of suche a man

3729. A. Owhar be þis ilche way?"
Cbg. Or here about by any way?
M. In þis ilche derne way?
Ca. Eyþir be paþ eyþir be way?
E(N). Here about by any way?
P. Lede a woman by any way?"

3730. A. "Sire, for gode," a seide, "nay!"
Cbg. "Sire," he sayde, "ffor gode nay!"
M. *like* A
Ca. "Sire," he sayde, "ffor gode nay!"
E(N). *like* A
P. "Syr," sayde the foster, "nay!"

3731. A. "What dones man ertow, bacheler?"
Cbg. What maner of man ertow here
M. What ertow, bacheler
Ca. What man ertow, he sayde, bacheler
E(N). What man ertow, bacheler
P. "What maner of man art thou, bachelere?

3732. A. "Sire," a seide, "a forster!"
Cbg. "Sire, I am a forster!"
M. *like* A
Ca. Sire, a seide, "I am a forster!"
E(N). *like* A
P. "Syr," he sayde, "I am a foster!"

3733. A. "Forster, so Crist þe be milde,
Cbg. Wylt þou, so Crist þe be milde
M. Wylt þou, so Crist be thy shild
Ca. Wylt þou, he sayde, so Crist be þe milde
E(N). So Crist þe fro shame shilde
P. "Wylt thou," sayde Beuys, "so Cryst the shylde,

3734. A. Wiltow lete cristen þis heþen childe?"
Cbg. Doo cristen þis heþen childe
M. Lete cristen me þis childe
Ca. Doo cristen here an heþen childe
E(N). Doo cristen þis ilke heþen childe
P. Do crysten here an hethen chylde?

3735. A. Riȝt, lo, now hit was ibore
Cbg. Riȝt, lo, now, hit was yn þe wode ibore
M. This same day it was borne,
Ca. For ryȝt now hit was ibore
E(N). Riȝt now ywys hit (he) was ibore
P. But ryght nowe was it borne,

3736. A. And ȝong hit haþ is moder for-lore:
Cbg. To ȝong hit haþ þe moder lore
M. To yonge the moder hath it lorne.
Ca. þe modyr it haþ to ȝung forlore
E(N). And (*om.* N) ȝong he haþ is moder lore
P. Ful erly it hathe the moder lorne,

3737. A. Wilt þow kep it for [to] min," a sede.
Cbg. *lacking*
M. And þow kep it to me, a sede
Ca. And kepen it wiþ mete and wede[21]
E(N). And kepen it wiþ mete and wede[21]
P. *see footnote below*[22]

3738. A. "And i schel quite wel þe mede?"
Cbg. *lacking*
M. I wyll the well yeld thy mede
Ca. And I schal quyte þe weel þy mede
E(N). *lacking*
P. *see footnote below*[22]

[21] In the Caius MS this line and its complement are placed after line 3734.

[22] For 3737–40 Pynson substitutes:

> And kepe it but this seuen yere,
> For X marke, and haue it here?"
> "Gladly!" sayde the foster tho;
> He toke the chylde & the golde also.

3739. A. þe forster him grauntede þer,
Cbg. þe forster grauntyd hym þer
M. *like* A
Ca. þe forster grauntyd Beues ryȝt þer
E(N). þe forster him grauntede þore (þer)
P. *see footnote below*[22]

3740. A. To kepe hit al þe seuen ȝer.
Cbg. For to kepe hit seuen ȝer
M. For to kepe hit seuen ȝer
Ca. To kepe þe chyld fulle seuen ȝer
E(N). For to kepe hem (him) seuen ȝer
P. *see footnote below*[22]

Detailed comment on the variations which occur in the above lines is not necessary. The question which I think we must ask ourselves in examining them is, which are merely scribal and which are not? The changes in the couplet of lines 3737–8 can hardly be due to scribal inadvertence. And if not, then what possible motive could lead two scribes copying from a written exemplar to change their copy (either or both of them) so that in one case we get "*God þe blesse, sire!*" *Beues sede* and in the other "*Crystys blessyng mote þe betyde*," especially as the change involves finding a new rime word? And the variant in the Pynson text not only makes use of a still different rime but so alters the idea that the thought of the couplet has to be completed in the subsequent line. All this seems to me to presuppose a scribe whose function is not to copy a text but to rewrite it, not in the interest of improving a line here and there but throughout the whole of a long poem, for the passage quoted can be matched almost anywhere in the romance.

If, on the other hand, the various forms of the passage reflect the recitations of minstrels who substitute and improvise when the memory fails to produce the exact words of a line or a couplet, and who for whatever reason are led to put down the story in writing, or assist in its recording, the differences in our existing texts are understandable.

What is true of the different manuscripts of *Beves of Hampton* is also true of *Guy of Warwick* and *Richard the Lion-Hearted*. But as I have explained above, the problem is complicated in the case of *Guy of Warwick* by the fact that the French original was turned into English, at least in part, four times, and in studying the variants we must be sure we are comparing passages of the same translation. I have already quoted a short passage from two manuscripts of one translation to illustrate the closeness of its dependence on the French. This passage would also serve my purpose here since it contains variations which do not seem to me to be scribal. But in order to present a further specimen for examination I select a passage from the second version in couplets, most completely preserved in the manuscript in the University Library, Cambridge, but some 2,000 lines of which are also represented in the Caius manuscript. And I also choose a short section which happens to be found as well in the brief fragment that has come down to us in Sloane MS 1044. In this way we may compare the readings of three manuscripts. Since the selection is quite short I include the French text and print the four passages in two sets of parallel columns:

Guy of Warwick

French ll. 7563 ff.		Caius MS ll. 7389 ff.
Ço fu en mai, el tens d'esté,		It was in may in somers tyde;
Que Gui ert en Warewic la cité.		Guy was at Warrewik in moche pride.
De berser est un jur repeiré,		From huntyng on a daye he is come,
Veneisun ad pris a grant plenté;		Grete plente of venyson he hath nome.
Mult joius e lé se feseit;	5	Moche ioye he made and solas,
A une vespree, que bele esteit,		So that in the euenyng so mery he was.
Gui en une tur munta,		
En halt as estres se pua;		
Le pais envirun ad esgardé		The Contree he behelde aboute farre,
E le ciel, qui tant ert esteillé,	10	And the skye thikke with sterre,
E le tens, qui ert serré e cler.		And the weder that was mery and bright.
Gui comence dunc a penser		Guy bethoughte him anone right
Cum Deus li out fait grant honur, . . .		That god him had so moche honour doo

Sloane MS 1044 (frag.) ptd. Zupitza, Wien. Sitzungs., **74**: 629

It was in may in someres tyde,
Guy was at Warwik wiþ pride.
From huntyng on a day was come,
Gode plente of venesoun had nome.
Muche joy he made and solace,
So þat in an evenyng, þat myry was,
Sir Guy to a toure steiȝ
And lened him to a corner an heiȝ.
He biheld þe cuntre about ferre,
Þe welkne, þat was wel þik of sterre,
And þe weder was myry and briȝt;
And Guy þouȝt him anon riȝt . . .
[The fragment breaks off at this point.]

Camb. Univ. MS ll. 7119 ff.

Hyt was in a somers tyde,
That Gye had moche pryde:
He came fro huntyng on a day
Wyth grete solace & mekyll play.
Þey toke plente of veneson
And broght hyt vnto the towne.
At euyn he wente into a towre
Wyth moche yoye and honowre.
He behelde there the ayre
And the lande, þat was so fayre.
The wedur was clere & sternes bryȝt.
Gye beganne to thynke ryght,
How god, that sate in trynyte,
Had made hym a man of grete poste

Here the Caius and the Sloane texts are fairly close to each other, and to the French; the differences can in most cases be explained as scribal variants. But the divergence of the Cambridge manuscript clearly goes beyond anything that we can attribute to scribal corruption. I have italicized the verses that are most conspicuously different, even where the rime words are the same. It will be noted that while the second line is basically the same in all three English texts it does not render *Warewic la cité* of the French.

In the case of *Richard the Lion-Hearted* the evidence of oral modification is not so frequent. The various manuscripts seem to reflect a written tradition to a greater degree than do the two romances previously examined. Nevertheless, oral modifications are at times evident. The manuscripts fall into two groups representing what the editor calls the *a* and the *b* tradition. Since we do not have the source of the romance for comparison, it is impossible to say whether or not the two traditions represent independent renderings of the assumed French original, but I have adopted the only safe procedure, that of examining the two as though they were independent versions. In his critical edition, Brunner prints as his basic text the Caius manuscript, representing the *a* tradition, supplying the gaps in this manuscript from the closely related text of Wynkyn de Worde. The variant which I print from B.M. Add. MS 31,042 belongs to the same group, and these are the only texts which preserve the *a* tradition. The verses which I believe show more than scribal variation are again italicized:

Wynkyn de Worde

The fyrste yere that he was kynge
At Salysbury he made a justynge
And comaunded euery man to be there,

Bothe with shelde and with spere,
Erles and barons euerychone;
At home ne dwelled neuer one.
On forfeyture, on lyfe and londe,
For nothynge that they ne wonde:
This was cryed, j vnderstonde,
Thorughout all Englonde.
All was for to loke and se
The knyghtes that best myght be.
There they came all at his wyll
His comaundemente to fulfyll.

B.M. Add. MS. 31,042

In þe fyrste yere that he was kynge
At Salysbery he made a grete justynge
And comanded þat euerilke knyghte
þat couthe hym were,
Oþer with shelde or with spere,
Erlles, barouns euerylkon
Lukes þat byhynd there byleue none
Ffor forfetoure of wyfe and lande
And for nothynge that they ne wande.
Þat was the crye j understande
Þat was cryed therowte all Yngland
And alle was to looke and to see
What knyghte beste doande myght bee
And thedir they come at þe kynges wyll
His commandement for to fulfyll.

The *b* tradition contains a 52-line passage not in *a*. Of this tradition the Auchinleck and the Douce manuscripts represent different branches, showing interesting divergences. The lines are inserted between 3346 and 3347 of the *a* tradition:

Auchinleck MS		Douce MS
Ffor wrath worth sike the Kyng of Ffraunce,	19	*Perfor euilyd þe Kyng of Ffraunce,*
His lechis seyde withoute distaunce	20	His leche seyde wiþowtyn distance
That he ne sholde neuer hol be		That he ne myth not hol bene
Bute he to Ffraunce tourne aye.		But he turnyd home ageyne.
The Kyng of Ffraunce tho vnderstode		*The kynges councel vnderstode*
That hure consail was trywe & goode		*& seyde þe leche was trewe and goode*
His shippes he dighte more and lasse	25	*He tok leue at Kynge Richard*
And wente hom atte Halwe-Masse.		*& at þe oder aftyrwarde.*
		The Duke of Burgonye in his stede he leet
		And of his ffolk a party gret.
Kyng Richard on hym gan crie,		Richard fast on hym gan crie,
And seyde he dude vileynye,		And seyde he dede gret vileynye
That he wolde for any maladie		That he wolde for maladie
Wende of the lond of Surrye,	30	Owte of the lond of Surrye
Er he hadde do Godes seruyse		*Wyl he wer jn Godes seruyse*
For lif or deth in any wise.		For lif or det or any wise.
The Kyng of Ffraunce wolde hym noght here,		*Kyng Philip nolde hym not here*
But wente forth on this manere.		*He partyde wrathe jn his manere.*
For they departede thus, for soth,	35	*And after þat partyng for soth,*
Euere after were they wroth.		Euere after þat þey were wroth.

Further comment on these passages is unnecessary.

I may close this discussion with a brief recapitulation of the argument and the conclusion to which I think it leads. The paper consists of six sections. In the first I outline the theory of oral composition as it has been developed in connection with the Homeric poems and applied to the Old French chansons de geste. According to this theory two basic practices are all-important to the singer of heroic poems who partly improvises, partly draws upon his memory: the extensive use (1) of *formulas* and (2) of *themes*. In the second and third sections I have shown that formulas and themes are present in certain Middle English romances to a degree fully as great as in the Greek epic and the French chanson de geste, and in the fourth section I have called attention to a feature of the Middle English romances—the predictable complement—which serves a similar purpose. Each poet or minstrel has to a certain extent his own favorite formulas and his own way of treating a theme, as would be expected, but many of these formulas and themes are the common property both of poets who compose in writing and (as I believe) of minstrels. Those who have developed the theory of oral composition have argued that the presence of these features in the *Iliad* or the *Chanson de Guillaume* justifies the conclusion that these poems were composed by those who recited them and in the process of recitation, just as the long narrative poems of Jugoslavia are composed by the *guslars* today. On this opinion I do not take a stand. I am inclined to think that it goes farther than is necessary, that it probably attaches too little importance to memory, and that, whatever changes a poem may undergo later, behind the poem stands the poet. This, I realize, would be criticized by proponents of the theory as begging the question, since in their view the reciter is the poet. However this may be, in the fifth section of my paper I show that the Middle English romances cannot in general have arisen in this way, since in so many cases they are translations of French romances, often so faithful to their source as to make it most likely that the English translator was working with his French original open before him. However, in the final section of my paper I have tried to show that improvisation has had a share in the production of the different versions of the same romance that have come down to us.

In examining the variations between manuscripts I do not suggest that we eliminate the scribe. Some manuscripts were copied and some copyists were careless. Not all the differences between one text and another reflect changes due to the reciter. I am only insisting that while some of these differences are such as can be accounted for by the familiar principles of hermeneutics, many others cannot. The variations when displayed in full, rather than obscured in a mass of textual notes, are seen to be of much the same kind as the variations between different

versions of a ballad. As for the ballad, most of us I suppose have given up the romantic idea that *das Volk dichtet*, but the late Professor Gerould once remarked that, if the people did not write the ballads, at least they rewrote them. I have elsewhere argued against the too easy acceptance of the view that the romances were written by minstrels, but in Professor Gerould's sense I believe they were rewritten by minstrels. Once we accept this possibility, many of the less admirable features of the romances—the many bad rimes, the clichés, the constant recurrence of ideas and expressions distinguished only by their lack of freshness, and the positive banalities—will appear not as an indictment of the poet but as the inevitable resort of one who in telling a story "cannot think without hurry about his next word, nor change what he has made, nor, before going on, read over what he has just written,"[23] but is under the immediate necessity of completing a verse, a couplet, or a rime. Whether the evidence which I have presented justifies this interpretation each one must decide for himself. If it does, then we may say that some of the Middle English romances owe their final form in part to improvisation.

APPENDIX

The following citations are intended to support the list (p. 426, above) of common elements which enter into the *theme* of a battle or an individual combat. Since there is no battle or combat in *Athelston*, there are no citations from this romance, and there are few extended combat-scenes in *King Horn* and *Havelok*. From the other romances the list could easily be extended and the examples under the various headings could be multiplied. It should be remembered that in the theory of oral improvisation the *theme* (and its components) does not call for similarity of wording.

1. *The opponent approaches.*

Brademond aȝenes him cam	Beves	994
King Grander was of herte grim & rod to Beues & he to him.	Beves	1745–6
Oȝaines sir Gij þer come Gayer, To iuste wiþ him he drouȝ him ner.	Guy	871–2
Þe douke come prikiand on his stede	Guy	899
He come als swiþe as he miȝt driue	Guy	907
Þan com prikeing dan Gwissard	Guy	1479
& Gij him was oȝain cominde	Guy	3508
Otus him went wiþ gret hete	Guy	5279
To the Kynge Rycharde gan hym dyght Than another noble knyght.	RLH	397–8

2. *The hero takes note, or "stood and beheld."*

Beues of him nam gode kep	Beves	998
Hauelok þat hauede spired wel Of here fare, euerilk del, . . .	Hav.	2620–1
Bradnond the kyng stode and beheld	Beves (Chetham)	762
Bothe partyes stode and beheld	*ib.*	777
Kyng Rycharde gan houe & abyde, Yf ony mo wold to hym ryde.	RLH	301–2
Kynge Rycharde houed and behelde	RLH	325
Styll he houed and bode þore	RLH	343
The aunterous þo turnyd agayn, And houyd stylle for to seyn. . . .	RLH	511–2
He houyd & byheeld vs ȝerne	RLH	559
Kyng Richard houyd & beheeld	RLH	3955
Sere Galabre houyd stylle To see who wolde ryde hym tylle.	RLH	5339–40

3. *He leaps into the saddle.*

And lep him self in to þe arsoun	Beves	876
Beues in þe sadel lep	Beves	989
And euery knyȝt lep on his stede	Beves	2328
And lopen to horse anon riȝte	Beves	4140
And lep to hors, wiþ outen oþ	Beves	4316
And lep vpon a sterne stede	Beves	4500
Þe stede toke bi the reyn, & lepe vp wiþ gret meyn	Guy	883–4
& dede him lepe opon his stede	Guy	3022
Vpon here steedes manly þey lepen	RLH	5111
Kyng Richard into þe sadyl leep	RLH	5741
And lopen on stedes sone anon	Hav.	2616

4. *He spurs his horse.*

Gij wiþ spors smot þe stede	Guy	1457
Wiþ scharp spors þai smiten her stede	Guy	5463

[23] Parry, *Harvard Stud. in Class. Philol.* **41**: 77, 1930.

He smot Arondel wiþ spures of golde Beves 999

They smote theire stedus with spurres of gold
Beves 775 (Chetham)

He smot Ffauel wiþ spores off golde RLH 5077

5. *He shouts an angry defiance or orders his opponent to yield.*

"Aȝilt þe, treitour, þow foule þef!
Þow hauest þe kinges sone islawe,
Þow schelt ben hanged & to-drawe!
Beues seide: "Be sein Ion,
Treitour was y neuer non:
Þat i schel keþe hastely,
Er þan ich wende, sikerly!" Beves 4374–80

Gij to smite he heyed bliue,
& seyd to him: "in iuel stounde
ȝaf þou þe douke Otous wounde.
To wroþer hele iuste þou wiþ him.
He is mi germain cosyn:
Icham þe douke Reyner þat to þe speke;
Icham y-comen him to awreke:
Turn þe and iuste wiþ me."
"Bleþeliche," quod Gij, "bi my leute." Guy 908–16

"Now ichil þe for a couward held,
& for a kniȝt vnwrast in feld:
Bot þou wilt wiþ me justi,
Ichil þe don a vilani.
Hennes forward war þe fro me,
Þi dedliche fo ichil now be!" Guy 1863–8

Herhaud him seyd, "Otus of Pauie,
Understond tow of þat felonie
Þat tow in Lombardi ous dedest,
When þou mi lord betreydest.
Wele we schul þer-of awreke be,
ȝif god wil, er þe sonne doun te."
Otus answerd, "þou lexst on me,
& þat y schal sone kiþe þe;
Gret scorn is here, so y go,
Y warn þe icham þi fo." Guy 2013–22

Whan Otes sawe sir Guyon
Comme rennyng to him as a lyoun,
With highe voice he gan vpbreide,
And to the Duke Otes thus he seide:
"Thou false and disceyuable traitour,
Full litell thou thoughte on thyn honour,
Whanne thow bee-traidest me,
And dud my men with sorowe slee
In the forest of playnes, as y forthe comme
Wiþ my felawes, good knyghtes echoone.
Fro hense forewarde, y telle the,
Thy dedly foo y shalbee.
In good poynte to bee y am not like,
Tille y haue thyn hede of strike." Guy 2071–84
(Caius)

"Esclandar," seyd Gij, "wende oȝain to me,
& forsoþe al siker þou be;
Drede þe of no noþer þan of me,
Ones to iusti ich oxi of þe."
Esclandar seyd, "artow Gij?
Ich þe defende sikerly.
Bi Mahoun þat ich leue opon,
Neuer no schal ich oway gon,
No neuer schal y bliþe be,
Til ich þat heued binim þe;
Bihoten ich it haue a maiden of pris,
Þe soudans douhter þat wel fair is." Guy 3041–52

Þan seyd þe treytour, "glotoun,
Dye þou schalt wiþ resoun:
Ich þe abie in þis stede."
Herhaud anon to him sede:
"Þou lexst," he seyd, "vile losaniour:
Þou it abist, bi seyn Sauour." Guy 5257–62

Þan seyd Gij to Otus so vnwrast:
"To me ward þou wende on hast,
& were þe of þat felonie
Þat þou dest me in Lombardye." Guy 5275–8

"Þou wroche glotoun losaniour,
Þou schalt þe ȝeld, bi seyn Sauour.
Now ichil ȝeld þe to prisoun
To mi lordes wille, þe douke Otoun."
"Þou lext," þan Gij sede,
"Þou liþer bodi, so god me rede.
Erst þou it schalt abigge,
Er þou me in prisoun legge." Guy 5317–24

(Afterwards, in all cases, they begin to fight fiercely.)

"Gij," quod he, "ȝeld þe anon!
ȝe ben ded now euerichon;
To þe douke we han trewþe pliȝt
To bring him þi bodi þis niȝt."
Wiþ þat ich word wel smert
Gij him smote vnto þe hert. . . . Guy 1361–66

"Gij," quod he, "ȝeld now þe!
It no may no noþer be:
On þe erþe liþe þi scheld to-dreued,
Nouȝt o pece is wiþ oþer bileued,
& þine helme is al to-hewe,
Þine hauberk to-rent þat was newe;
& wounded þou art, þou miȝt well se,
Long miȝt tow nouȝt oliues be.
To day ichil ȝeld þe to þe douk Otoun,
& he þe schal do in his prisoun.
Þan seyd Gij, "Gwichard, y nille:
To ȝeld me to þe is nouȝt mi wille,
Þer-wiles ichaue mi swerd y-grounde,
& mi bodi wiþouten wounde." Guy 1481–94

"Come ffyȝt wiþ me now hoo þat dar!" RLH 4538

Þe Sarezynys cryyd in here langage:
"Crystene houndes off euyl rage,
But ȝe the sooner wende home,
Here ȝe haue ffet ȝoure dome!"
Whenne Kyng Richard herde þat cry,
He swor hys oþ be Seynte Mary
Þe Sarezynes scholde be hangyd alle
Or swylke a cas hem scholde befalle. RLH 6074–80

6. *He rides against the foe.*

To hem faste he gan ride Beves 881

And Saber rod him aȝene Beves 3400

He rod to him wiþ gret randoun Beves 3809

To king Yuor he gan areche Beves 4189

And rod forþ wiþ gret randoun Beves 4501

Oȝain to þe douke he is y-come Guy 924

Toward Hugoun he made a ras Guy 1428

Forþ þai ȝede wiþ gode welle Guy 5115

Gij him wiþ-went, wold he nouȝt fle Guy 6971

And a spere to hym he bare RLH 318

Towarde hym he gan ryde RLH 348

He prycked forth vpon his stede RLH 2289

Sere Archade took a gret launse,
And come prykande wiþ bobaunce. RLH 4521–2

Agayn Sere Archolyn hym dyȝt RLH 5304

7. *The knights meet head-on.*

And ase þei boþe to gedre mete,
Wiþ here launces þei gonne mete,
Þat hit gonnen al to-driue
& to-borsten on pises fiue. Beves 1747–50

Wiþ þat hii ride to gedres boþe,
Ase men þat were in hertte wroþe,
So harde þai gonne to gedres mete
And wiþ here launces gonne grete,
Þat þourȝ þe scheldes þe speres ȝode; Beves 4151–5

They smytten than to geders thoo,
Full harde strokes they yiuen both two. Guy 875–6 (Caius)

To-gider so hard gyn þai driue,
Þat her speres gan al to-riue. Guy 943–4

Togider þai smiten wiþ gode wille,
Þat boþe of her hors adoun felle; Guy 2023–4

With that either of theim pricked his stede,
And in grete wrathe to-gider yede. Guy 2085–6 (Caius)

Sone there beganne a straunge shoure:
To-geder they smyte knyghtis of valoure. *ib.* 2181–2

Al togider þai gun smite;
Semblant of loue þai kidde bot lite. Guy 2213–4

Her steden þai turned snelle,
& to-gider þai smiten wiþ gode wille. Guy 3053–4

Heteliche togider þai smete
Opon her scheldes wiþ gret hete. Guy 6973–4

And smot togedere wiþ grete randoun RLH 4845

Wiþ egyr jre togedere rode,
Þat eyþere stede to grounde glode,
And brak here nekkes in þat stounde,
Þat þey lay ded vpon þe grounde. RLH 5347–50

Þey reden togedere, as we rede,
Þat boþe to þe erþe þey ȝede. RLH 5319–20

8. *They strike each other.*

Eche man on other faste gan smyte Beves 3714 (Chetham)

Ich smit on oþer wel gode dint;
Þai smiten togider for soþ, y pliȝt,
Eueriche to nim oþer dede his miȝt. Guy 886–8

To Gij he smot wiþ gret hete,
& Gij oȝain to him smite. Guy 941–2

Þai smiten togider hard & wel
Wiþ her swerdes of grounden stiel Guy 945–6

& after þai drouȝ her swerdes newe,
Wiþ gret envie to-gider þai hewe. Guy 2025–6

Than gan they to-geder smyte:
Noon spared other bot a lite. Guy (Caius) 2037–8

To-gider they smyte than faste:
Of the Almaignes they were not agaste. Guy 2123–4 (Caius)

Togider þai smiten hard and wel
Wiþ brondes wele wrouȝt of stiel. Guy 2227–8

Seþþen þai drouȝ her brondes of stiel,
& smiten togider hard & wel. Guy 3003–4

Now þai ginne togider smite:
Non no spared oþer bot lite. Guy 5135–6

Togider þai smite hard wiþ alle,
Mani on þer was ded & doun falle. Guy 5137–8

To-gider fast þai gun smite Guy 5197

And smot togedere wiþ grete raundoun RLH 4845

Þey smeten togedere dyntys sare RLH 5305

And smot Godrich, and Godrich him,
Hetelike with herte grim. Hav. 2654–5

Þer mouhte men se two knihtes bete
Ayþer on oþer dintes grete. Hav. 2664–5

9. *Fire flies.*

Þe fur fleȝ out, so spark o flintes — Beves 1754

Þe fure brast out so brond ibrent — Beves 4165

Þe fur sprong out after þe dent — Beves 4516

Þat fire vnder þe fet aros — Guy 1459

Of that dente the fyre outspronge — RLH 354

10. *One strikes the other's shield.*

Ayther smote oþure in myddus the sheld — Beves (Chetham) 778

Beues spere held and he smote faste,
That thorouȝe sheld and actoun
He bare thorough knyght Radison. — Beves 780–2 (Chetham)

With þat word Beues smot doun
Grander is scheld wiþ is fachoun. — Beves 1767–8

And wiþ here launces gonne grete,
Þat þourȝ þe scheldes þe speres ȝode;
At þe breinies þe dent wiþ-stode. — Beves 4154–6

Gaier smot sir Gij bifore
& þurch þe scheld him haþ i-bore. — Guy 877–8

He smot þe douke on þe scheld,
Þat it fleye in þe feld. — Guy 919–20

Sir Torold smot him on þe scheld,
Þat he feld him in the feld. — Guy 955–6

Wiþ al his miȝt he smot him to,
Wel euen he clef his scheld þo. — Guy 1461–2

Otes smote Guy in the sheelde — Guy 2087

Ermine he smot on þurch þe scheld;
Almost he feld him in þe feld. — Guy 3499–3500

He smot Gij on þe scheld bifore,
Þat neye he haþ his swerd forlore. — Guy 4777–8

He smot him ouer þe gilden scheld — Guy 5217

Heteliche togider þai smete
Opon her scheldes wiþ gret hete. — Guy 6973–4

Full egerly he hym grette
With a dente amyd the shelde. — RLH 290–1

Swylke a strok on his scheld he set — RLH 478

The auenturous smote his shelde amydde — RLH 495

He hytte hym vpon þe scheelde — RLH 4526

And in hys scheeld he smot hym soo,
Þat it cleuyd euene in twoo. — RLH 4543–4

He smot an amyral in þe scheelde — RLH 5080

11. *Part of the shield (helmet, hauberk) flies off.*

A quarter fel in to þe feld. — Beves 1760

Þat his scheld fleȝ fram him þore — Beves 1905

And Yuore wiþ þe strok of yre
Made fle in to þe riuere
A large quarter of his scheld. — Beves 4199–4201

He smot þe douke on þe scheld,
Þat it fleye in þe feld. — Guy 919–20

On þe erþe liþe þi scheld to-dreued — Guy 1483

Sadok toforn haþ him smete
Of his scheld a quarter wiþ gret hete. — Guy 1871–2

Otes smote Guy in the sheelde,
That euen half flowe in the felde. — Guy 2087–8

He bare awaye halfe his schelde — RLH 320

Hys schelde in twoo peses roff — RLH 506

Smot Sere Thomas, wiþouten lye,
On his spawdeler off his scheelde,
Þat it fleyȝ into þe ffeelde. — RLH 5326–8

Þat a quarter [*of the helmet*] out fleye — Guy 1497

Of her helmes þe flours gan fle — Guy 3007

A quarter of his helme doun riȝt — Guy 5256

& of his helme he carf a quarter — Guy 5286

A quarter of his helme he smot adoun — Guy 5390

Þat sercle of gold & is crestel
Fer in to þe mede fel. — Beves 4175–6

Of his hauberk it tok a pece brod — Guy 1500

12. *The hand goes too.*

And is left hande be þe wrest,
Hit fleȝ awei þourȝ help of Crist. — Beves 1769–70

And his left hond, be þe wrest
Hit fleȝ awei þourȝ help of Crist. — Beves 4207–8

Þat he made upon the feld
His lift arm fleye, with the swerd. — Hav. 2634–5

His riȝt arm & is scholder bon
He made fle to gronde anon. — Beves 4217–8

For his swerd he hof up heye,
And þe hand he dide of-fleye. — Hav. 2750–1

13. *He strikes the helmet.*

And sum vpon þe helm a hitte — Beves 865

Wiþ outen eni wordes mo
Beues Brademond hitte so
Vpon is helm. . . . — Beves 1037–9

Doun of þe helm þe swerd gan glace
And karf riȝt doun before is face. Beves 4177–8

Beues a hite on þe helm of stel Beves 4504

Gwichard smot Gij wiþ michel miȝt
Opon þe helme þat schon so briȝt. Guy 1495–6

So strong is þat strok y-ȝiue,
Þat his helme is al to-driue. Guy 1875–6

Ermine smot him on þe helme an heyȝe Guy 3503

On heye on helmes he hem smot Guy 3555

He smot him on þe helme briȝt Guy 5255

On þe helme þat strok glod Guy 5265

Of þat helme þat swerd glod Guy 5289

Opon his helme he hewe gode won Guy 5306

& on his helme hard him clout Guy 5408

Þai hewe on his helme, þat blod out ran Guy 5409

A stroke he thought to be set well
On his helme that was so stronge. RLH 352–3

He set his stroke on his yron hat RLH 367

He smote hym on his bassenet RLH 403

There he gaue the kynge his dele
That his helme al toroue. . . . RLH 416–7

14. *He strikes the head.*

Sum he smot opon þe hode Guy 1397

Opon þe heued Gij him smot Guy 4769

Hyȝ mase vpon his heed he layde RLH 412

He smot him on þe heued so Hav. 2642

And smot him so upon þe crune Hav. 2734

15. *He strikes the head off, or it flies off.*

Þat þe heued fleȝ in to þe riuere Beves 638

Men myght se ouer all
Hedys cirlyng as a ball. Beves (Chetham) 799–800

And sum kniȝt Beues so ofrauȝte,
Þe heued of at þe ferste drauȝte. Beves 867–8

And smot of king Grander is heued Beves 1776

And tok him to þe heued anon
And strok hit fro þe scholder bon. Beves 4237–8

Þat his heued fram þe bodi flei Guy 1383

His heued of fleye wiþ þat drauȝt Guy 1478

Þat heued fram þe bodi he slouȝ Guy 2940

Þe heued of a Sarrazin he dede of fle Guy 3018

His heued he dede fleye in þe feld Guy 4774

Þe heued he dede fram þe bodi fle Guy 5402

Þat helm & hed ffleyȝ into þe ffeeld RLH 3156

Þat þe hed trendelyd off as a bal RLH 4534

Men of armes þe swerdes out breyde,
Balles out off hoodes soone þei pleyde. RLH 4551–2

Þat on þe schuldre ffel þe brayn RLH 4555

And of som he pared so þe croune,
Þat helme and hed fel adoune. RLH 5095–6

Þat helme & hed ffleyȝ into þe ffeeld RLH 5852

At eureche dunte
Þe heued of wente. Horn 609–10

Þat his heued of he plette Hav. 2626

And his heued al of-slawen Hav. 2676

16. *He strikes the neck, shoulder, side, etc.*

And sum he strok of þe swire Beves 637

And smot ato his nekke bon Beves 1919

Þat his nek-bon brac atvo Guy 5438

His necke he brake there a two RLH 298

Þat he smot hise necke on to Hav. 1823

In þe nekke he hytte hym wiþal RLH 4533

Vpon þe scholder in þat tide Beves 4191

Gij þurch þe scholder him smot Guy 903

Opon his scholder þat swerd glod Guy 1499

Gij he smot in þat stounde
Þurch þe scholder a grete wounde. Guy 5303–4

Opon his scholder þat swerd glod Guy 5391

Þurch þe scholder wiþ a swerd in Guy 5416

& þurch þe scholder he ȝaf him wounde Guy 6465

Some he smote on þe schuldyr-bon RLH 5093

And þat [he] bi þe shuldre-blade
Þe sharpe swerd let [dune] wade. Hav. 2644–5

And smot him on þe sholdre so Hav. 2738

And sum he smot þurch þe side Guy 1399

He bar hym þorwȝ vndyr þe scheeld RLH 5772

He smot him þureȝ þe herte Horn 875

17. *The spear breaks.*

Sir Radisons spere in sonder braste	Beves 779 (Chetham)
The spere brast, the knyght fell doun	*ib.* 783
& to-borsten on pises fiue	Beves 1750
Þat his spere al to-fliȝ	Beves 2536
Þe launce brak, þat was wele wrouȝt	Guy 879
Þat her speres gan al to-riue	Guy 944
Þe launse barst, þe Sarezyn sterff	RLH 5090

18. *He unhorses his opponent.*

A bar him doun of his hors	Beves 3406
Kniȝtes out of sadel iboren	Beves 3803
& felde him ouer is horses croupe	Beves 3818
& bar þe douke Reyner saunfeil Ouer & ouer his hors tayl.	Guy 921–2
And oute of his sadell he did him bere	Guy 2130 (Caius)
Þat of his hors he felled him tho	Guy 2134 (Caius)
That of his stede he felled him downe	Guy 2177 (Caius)
Of his hors he has him feld	Guy 4773
Of his stede he feld him, wiþouten fable	Guy 5090
Of his hors he haþ him feld	Guy 5125
Of his stede he feld him doun riȝt	Guy 5134
Of his hors he feld him wiþ hete	Guy 5148
Otus he smot þer-wiþ so, Þat he les his stiropes to.	Guy 5273–4
The aunterous felde hym þere wiþ yre Doun off his stede, & brak hys swyre.	RLH 497–8
And bar hym ouer hys hors-bak	RLH 508
Doun off hys hors almost he fleyȝ	RLH 546
Off hys hors top ouer taylle	RLH 4564

19. *He fells horse and man.*

That hors and man he ffellid to ground	Beves (Chetham) 814
Euer alse he com þan, And felde boþe hors & man.	Beves 3417–8
Boþe man & hors he slouȝ	Beves 3428
Þat man and hors fel ded to grounde	Beves 3434
Þat ded he feld boþe hors & man	Guy 5480
Þat hors & man it fel to grounde	Guy 6466
His horse and he fell to grounde	RLH 299
Þat hors & man ouerþrewe	RLH 479
Þat man and hors he cleuede doun	Beves 4514
Hors and man he felde to grounde	RLH 4894
And al tobarste hors & man	RLH 5137
Hors & man to grounde ȝode	RLH 5788
Þat hors & man ffyl ded to grounde	RLH 7137

20. *Both are unhorsed.*

So harde þai þreste to gedre þo, Þat here gerþes borste ato, & felle to grounde boþe þo, A fote nedes þai moste go.	Beves 4157–60
Þat boþe of her hors adoun felle	Guy 2024
Þat of her hors þai fellen bo	Guy 3002
And he ouer his sadell droue	RLH 418
So þat he boþe felle dune, To þe erþe, first þe croune.	Hav. 2656–7

21. *He draws his sword.*

Anon riȝt is swerd out drouȝ	Beves 274
Þo Beues hadde þat swerd an hond	Beves 863
Here swerdes drowe kniȝtes stoute	Beves 1751
& drouȝ his swerd anon riȝte	Beves 2542
& after þai drouȝ her swerdes newe	Guy 2025
And than he drowe his swerde browne	Guy 2178 (Caius)
His swerd of stiel he haþ up pliȝt	Guy 2401
Sir Gij his gode swerde out drouȝ	Guy 2939
Seþþen þai drouȝ her brondes of stiel	Guy 3003
His gode brond þan drouȝ he	Guy 3017
Ac hetelich he pliȝt out þat brond	Guy 5421
Swerdes þai drowen, wiþouten feyle	Guy 5481
Anon his swerd he haþ out drawe	Guy 6469
Men of armes þe swerdes out breyde	RLH 4551
Hastely swerdes þey drowen	RLH 4829
In his hond his swerd ut-drawe	Hav. 1802
Þe seuenþe brayd ut his swerd	Hav. 1825
Þey drowen ut swerdes, ful god won	Hav. 1837

William Wendut his swerd vt-drow Hav. 2632

Huwe Rauen ne forgat nougt
Þe swerd he hauede þider brouht;
He kipte it up, and smot ful sore. Hav. 2636–8

Grundlike here swerdes [he] ut-drowen Hav. 2659

He grop þe swerd ut sone anon Hav. 2728

22. *The sword bites well.*

With Morgelay, þat wel wold byte Beves 2451

Wiþ Morgelai, þat biter bot Beves 2492

With Morgelay, þat wolde wel bite Beves 3407

Wiþ his swerd þat wold wele bite Guy 1506

Wiþ her swerdes þat wil wel bite Guy 2934

Wiþ his fauchon þat wele bot Guy 3556

Wiþ his swerd þat wele bot Guy 3562

Wiþ swerdes þat wil wele bite Guy 5198

Swerd he bar þat wele wald bite Guy 5793

Wiþ hys fawchoun, þat byttyr bot RLH 4874

23. *He strikes with sword.*

Wiþ Morgelay a smot him þo Beves 4205

And wiþ his swerd he smot him so Guy 1437

Gij þat on wiþ his swerd rauȝt Guy 1477

& wiþ his brond he him smot Guy 5380

Wiþ swerdes þai smiten him þan about Guy 5407

He layde on wiþ his sworde RLH 5065

With þe swerd so he him grette Hav. 2625

24. *They lay on eagerly, fight "as they were wod," spare not.*

Þo laide þai on wiþ eger mod Beves 1009

Togider þai smiten wiþ gode wille Guy 2023

Al togider þai gun smite;
Semblant of loue þai kidde bot lite. Guy 2213–4

Þer þai fouȝten togider snelle,
& smiten strokes wiþ gode wille. Guy 5381–2

Þai hewe and slouȝ wiþ gret hete Guy 5483

And slowe Sarsins, as hii wer wod Beves 1010

He fauȝt ase he wer wode þer fore Beves 1772

Þai leide on ase hii were wode Beves 3799

He fauȝt ase he wer wod þer fore Beves 4210

And layden on as þey were woode RLH 5105

And stirte forth so he were wode Hav. 2609

And fouhten so þei woren wode Hav. 2661

He nold spare him bot lite Guy 918

No spard he for no drede Guy 1367

Herhaud smot anoþer and spared nouȝt Guy 1376

& nothing he no spared him Guy 1870

Noon spared other bot a lite Guy (Caius) 2038

Spare nolde [he] neuer on Guy 4772

Gij smot Hogoun, and nouȝt him for-bare Guy 4780

Non no spared oþer bot lite Guy 5136

Aiþer no spard oþer no-wiȝt Guy 5386

For was þer spared heued non Hav. 1906

For-bar he noþer kniht ne sweyn Hav. 2623

25. *He cleaves opponent to the saddle, chin, etc.*

In to þe sadel he hem slitte Beves 866

And some to þe sadyl he slyt RLH 3154

Into þe sadyl he cleff þe fferþe RLH 3159

And karff hym to þe sadyl anon RLH 5094

At þe girdel þe swerd astode Guy 1398

Vnto þe girdel stede þat swerd bot Guy 4770

Þat he cleue him to þe teþ Guy 3504

Heuedes cleue vnto þe chinne Guy 1394

Þe þrytteneþe to þe chyn he kerff RLH 5089

26. *He cuts the body, etc., in two.*

Þat is bodi a clef ato Beves 874

That all he hitt he cleve in two Beves 3716 (Chetham)

Þe bodi atvo he haþ to-deled Guy 2957

& to-clef him wiþouten abod Guy 4782

His bodi he clef adoun riȝt Guy 5120

Þat his heued he him to-clef Guy 6905

Þat he þe heued clef a-two Hav. 2643

Þat his scheld he clef ato — Beves 4206

Hys schelde in twoo þeses roff — RLH 506

So þat he clef his sheld on-two — Hav. 2730

27. *He kills opponent or gives a deadly wound.*

Saber slouȝ him in þat stounde — Beves 3402

Gij him smot vn-to þe hert — Guy 1366

Þat ded he feld him in þe mede — Guy 1368

Sleyne he hath the good Gayer — Guy (Caius) 2172

Dan Gauter he haþ y-slawe,
& gode Gilmin his felawe. — Guy 2989–90

Þe þridde to deþ he dede do — Guy 3020

And wiþ þat launse a kyng he slowȝ — RLH 5084

Til þat he hauede anoþer slawen — Hav. 2630

And þe þredde so sore he slow — Hav. 2633

Þar wiþ a ȝaf hem dedli wonde — Beves 636

28. *He pierces him through the body, heart, etc.*

Þourȝ is bodi wente þe dent — Beves 4383

Þurch þe bodi þe swerd gan gon — Guy 1372

Þurch þe bodi his swerd glod — Guy 1377

Þat þurch his bodi þe swerd gan go — Guy 1388

Þurch þe body he him smot — Guy 1431

Þurch his bodi þat swerd ȝede — Guy 1439

Þurch his bodi þe swerd he þriste — Guy 1463

Þurch þe bodi ful ney þe hert — Guy 1508

& þurch þe bodi he Sadok smot — Guy 1879

Þurch þe body he ȝaf him wounde — Guy 2937

Þurch þe bodi he him smot — Guy 2945

Þurch his bodi þe launce glod — Guy 2951

Þurch þe bodi he dede þe launce go — Guy 2980

Þurch þe bodi [he smot him] wiþ his launce — Guy 3060

Þurch his bodi þe swerd ȝede — Guy 5131

Þurch þe bodi he smot him anon — Guy 6427

Þurch his bodi þe swerd is gon — Guy 6965

Þurch out his hert þe launce he bar — Guy 1415

Þe dynt smot þorwȝ þe heþene herte — RLH 5081

Þe sharpe swerd let [dune] wade
Þorw the brest unto þe herte. — Hav. 2645–6

Thurgh his theighe he thruste his swerde grounde — Guy (Caius) 2090

Þurch þe thei Gij he haþ y-smite — Guy 5280

Thurgh-oute his hepe it glode — Guy (Caius) 2144

Þo yaf Godrich þorw þe side
Vbbe a wunde ful un-ride — Hav. 2672–3

29. *He fells him to the ground.*

Þe ten forsters wer feld te grounde — Beves 886

Sexti þosent þai felde to grounde — Beves 1018

Þat a felde him flat to grounde — Beves 1040

An hunderid Sarzins ffellid to ground — Beves 794 (Chetham)

Þe prince a felde in þe feld — Beves 3811

For smertte Yuor in þat stounde
Fels a knes vnto þe grounde — Beves 4193–4

To ground he felde him fot hot — Guy 882

Sat he neuer so wel no so fast,
Þat he no feld him sone on hast. — Guy 893–4

& feld him to grounde fot hot — Guy 904

& of his hors him haþ y-feld
Riȝt long streȝt in þe feld. — Guy 951–2

Þat he feld him in the feld — Guy 956

Þat ded he feld him in þe mede — Guy 1368

Þat ded he feld him in þe way — Guy 1390

Adoun he feld him ded riȝt þar — Guy 1416

In þe sond he feld him doun — Guy 1465

And of his stede he felled him ther — Guy 2128 (Caius)

That dede he did him to grounde goon — Guy 2164 (Caius)

& dede he feld him on þe grounde — Guy 2938

Ded he feld him an hast þo — Guy 2994

& adoun feld in þe feld — Guy 2996

Almost he feld him in þe feld — Guy 3500

Al ded he made him on þe grounde to lie — Guy 3505

Þat he feld him anon fot hot — Guy 3510

Þat ded he feld him anon riȝt — Guy 3516

Mani on he feld ded anon riȝt — Guy 5108

Þat ded he feld him on þe ston — Guy 5122

In þe sond he haþ him slawe Guy 5128

Ded he feld him of his stede Guy 5132

Herhaud him haþ þer afeld,
Þat dede he lay in þe feld. Guy 5229–30

Anoþer he feld þer anon Guy 6966

Þat he haþ þe grounde y-rauȝt Guy 6978

Hauede [he] ben brouht to grunde Hav. 2675

And smot him so up-on þe crune,
Þat Godrich fel to þe erþe adune. Hav. 2734–5

On knees he fyl doun and cryde "Creaunt!"
RLH 5361

Behynde hym þe spere out wente.
He leet hym lye vpon þe grene. RLH 5780–1

30. *The enemy falls to the ground (dead).*

Þat he ne fel ded to þe grounde Beves 1005

And smot ato his nekke bon:
Þe geaunt fel to grounde anon. Beves 1919–20

Ded a fel on þe pauiment Beves 4384

Ded he fel wiþ-outen abod Guy 1378

Of his hors ded he gan falle Guy 1434

Þat ded he fel doun fot hot Guy 1880

Þat ded he fel anon riȝt Guy 2232

Ded he fel wiþ-outen abod Guy 2952

Þat he fel doun in þe feld Guy 2958

Þat ded he fel wiþouten letting Guy 3522

Mani on þer was ded & doun falle Guy 5138

And the knyght fell to grounde
Full nye deed in that stounde. RLH 293–4

He fel doun ded as ony ston RLH 798

And he fell downe deed in haste RLH 2013

So þat he stan-ded fel þor dune Hav. 1815

Þat þe erl fel dun a-non,
Al so ded so ani ston. Hav. 2648–9

31. *Armor availed not.*

Hauberk ne scheld ne actoun
Ne vailede him nouȝt worþ a botoun Beves 1003–4

Þar nas non armur in þat londe,
Þat miȝte þe geauntes strok astonde. Beves 3429–30

His armour no was him worþ a slo Guy 1414

His sheld nas nouȝt worþ a botoun,
No his tvifold armes halp him nouȝt
Þat in Loreyn weren y-wrouȝt. Guy 2216–8

Scheld no hauberk nas him worþ a slo Guy 2936

Al was nouȝt worþ he had opan Guy 2950

Scheld no hauberk halp him no wiȝt Guy 3057

His hauberk nas him worþ a slo Guy 5076

His armes was him nouȝt worþ Guy 5130

Nouȝt worþ a botoun it [brini] nas Guy 6460

Ffor non armour wiþstood hys ax RLH 3163

Þer was non armure, verraymente,
So good þat myȝte wiþstande his dent. RLH 4869–70

Non armure iwrouȝt wiþ hande
Myȝte Kyng Richardes ax wiþstande RLH 5097–8

None off hys armes myȝte laste RLH 5773

32. *He taunts the vanquished foe.*

"Reste þe," queþ Beues, "heþen hounde!
Þe hadde beter atom þan here!" Beves 1006–7

(Yvor calls on his gods):
Þo Beues herde him so grede,
He seide: "Yuor, let be þat cri
And clepe to god and to Mari,
And let þe cristen, er þe deie,
Or þow schelt go þe worsse weie
And wiþ outen ende dwelle
In þe stronge peine of helle!" Beves 4224–30

A sede anon after þat dint:
"Treitour! now is þe lif itint:
Þus men schel teche file glotouns,
Þat wile misaie gode barouns!" Beves 4385–8

"Bi þe trewþe y schal mi leman ȝeld,
To day no schaltow þi trewþe held!" Guy 1369–70

"No þou, treytour, no schalt me lede
To þe douke that is ful of qued;
To his presoun no worþ y for þe brouȝt."
Guy 1373–5

And sayde: "Dogge, þere þou ly,
And reste þe þere tyl domysday,
Ffor þou art payyde off þy pay!" RLH 4546–8

33. "*With that came. . . .*"

Wiþ þat com þe bor aȝen Beves 806

Wiþ þat was come king Yuore Beves 3967

Wiþ þat com Beues in þat tide Beves 3971

Wiþ þat come Herhaud priking Guy 949

Wiþ þat come prikeing sir Urri Guy 958

Wiþ þat come a Lombard ride Guy 1359
Þan com Torald, a gode kniȝt Guy 1379
Wiþ þat come Urri prikeinde Guy 1385
Wiþ þat come ride þerl Lambard Guy 1407
Wiþ þat him come forþ Hougoun Guy 1417
Wiþ þat com his folk prikeinde Guy 2030
Wiþ þat come prikeand Tirri Guy 2229
Wiþ þat come Tebaud prikeinde Guy 2947
Wiþ þat come sir Morgadour Guy 2959
Wiþ þat come Esclandar prikeinde Guy 2975
Wiþ þat come forþ an amireld Guy 2987
Þan come Auþer ouer þuert Guy 3501
Wiþ þat come þe king of Nubie Guy 3506
Wiþ þat come þe gode Hogoun Guy 4775
Wiþ þat com forþ sir Gij Guy 5117
Þan come þer bi an hongend hille
Þe miȝti and þe hardi Gyoun Guy 5270–1
Wiþ þat þer com a stout Lombard Guy 5301
Wiþ þat come Herhaud prikeinde Guy 5337
Wiþ þat come þerl prikeinde Guy 6967
Wiþ þat fiue dukes off Heþenys
Come . . . RLH 7141–2

34. *He smote him.*

Þe kinges stiward a hitte so . . . Beves 873
Beues spere held and he smot faste Beves 780 (Chetham)
To Beues he smot a dent ful sore Beves 4174
Gaier smot sir Gij bifore Guy 877
Gij afterward Gaier smot Guy 881
Ich smit on oþer wel gode dint Guy 886
Gij him turned & gan to smite Guy 917
To Gij he smot wiþ gret hete Guy 941
Þai smiten togider hard & wel Guy 945
Herhaud smot anoþer and spared nouȝt Guy 1376
& so wele his stroke he sett Guy 1382
A Lombard he smot þo Guy 1387
So he smot him, for soþ to say Guy 1389
Wiþ swiche hete he smot him to Guy 1413
Wiþ al his miȝt he smot him to Guy 1461
Gwichard smot Gij wiþ michel miȝt Guy 1495
Gwichard he wald fond to smite Guy 1505
To him he smot swiþe smert Guy 1507
Now Sadok smot to Segyn Guy 1869
Heteliche to him smot Gyoun Guy 2215
Wiþ him smot þe douke Segyn Guy 2225
Otus he smot þer-wiþ so Guy 5273
Wiþ gret hete he smot Gyoun Guy 5281
Gij he smot in þat stounde Guy 5303
On of her kniȝtes he smot so Guy 6964
To þerl Florentin þan smot he . Guy 6972
A grete stroke he hym bare RLH 296
And smote to hym with jrefull mode RLH 366
To hym a stroke he dyght RLH 401
And another stroke he hym brayde RLH 411
He kipte it up, and smot ful sore Hav. 2638

35. *The victor returns home, or departs, with or without prisoners.*

So opon þe þridde day
Þe kniȝtes tok her leue and went oway. Guy 971–2
Wiþ þat he is out of the place y-went Guy 1881
& þe oþer turnen oȝen Guy 2248
Gij & Tirri oȝain ben y-gon:
Wiþ hem þai ladde her prisouns ichon. Guy 5155–6
Gij, & Tirri, and Herhaud also
Oȝain to þe cite þai ben y-go. Guy 5501–2

36. *The knight is stung by injury or anger to retaliate.*

Þo his bodi began to smerte
He gan plokken vp is hertte. Beves 631–2
King Grander was agremed strong,
Þat sire Beues him stod so long. Beves 1757–8
Beues þer of was agreued
And smot of king Grander is heued. Beves 1775–6
"O," queþ Beues, "so god me spede,
Þow hauest don gret vileinie . . .
Bo god, i swere þe an oþ:
Þow schelt nouȝt, whan we te-goþ,
Lauȝande me wende fram,
Now þow hauest mad me gram!" Beves 1890–98

Þo wex Beues in gret grame
And þouȝte wel wiþ Morgelay
Ȝelden his strok, ȝif þat he may. Beves 4186–8

Ac vp he sterte in haste þan
& in wraþþe to Beues ran. Beves 4195–6

Þe douke Reyner seye þat cas
Of Sessoine: wel modi he was. Guy 905–6

It oþouȝt Gij þo he was war;
Wiþ þerl Lambard he wald iusti,
& awreke þe gode Vrri. Guy 1410–2

For his deþ he was sori:
Him to awreke he haþ gret hy.
Neuer ȝete so sori he no was. Guy 1425–7

When Gij seye Herhaud y-feld,
To-hewen his hauberk & his scheld
(& of his hors feld he was,
As ded man lay on þe gras;
He seye þe blod þat cam him fro),
Wonder him þouȝt, & seyd þo:
"Þou lording, to þe y sigge
His deþ þou schalt wel sore abigge!
So mot ich euer word speke,
Mi maisters deþ ichil awreke,
& for a couward ich held þe:
Þou slouȝ him, & lete me be.
Bi him þat made sonne & mone,
Þou schalt it wite swiþe sone,
Þat tow schalt it biȝelp nouȝt
Þat he is to deþ y-brouȝt." Guy 1441–56

When Gij seye him so smite,
He was wroþ, ȝe may wele wite. Guy 1503–4

Þe douke him wreþþed for þat smite,
& was ful wroþ, ȝe mow wele wite. Guy 1877–8

Sleyne he hath the good Gayer:
In fraunce he was borne, Guyes feere;
To Guy he was leef and dere.
Whan Guy that sawe he was sory:
To the Duke he smot greuously. Guy 2172–6 (Caius)

So sone so douke Segyn seþ þis,
Wel wroþ he was wiþ him y-wis. Guy 2233–4

Sone so Esclandar y-seye þis,
To awreke þe amiral lef him is. Guy 2997–8

When þe douke of Tire þat y-seþ,
His men dye on so reweliche deþ . . .
He forþ ȝede . . . Guy 3511–15

For schame he hadde of þe stroke þare
Gij smot Hogoun, and nouȝt him for-bare. Guy 4779–80

To awreke him he hadde gret hye Guy 5284

Kyng Richard was agreuyd sore
Ffor þe strok þat he hadde þore. RLH 5309–10

Hwan Hauelok saw his folk so brittene,
And his ferd so swiþe littene
He cam driuende up-on a stede,
And bigan til him to grede,
And seyde . . . Hav. 2700–4

37. *He is "full woe," "in heart wroth," etc.*

Þar fore was Beues swiþe wo Beves 3414

In his hert him was ful wo Guy 2992

Sir Gij of-toke him nouȝt, þerfor wo him was Guy 3062

Than the emperour was full wo RLH 2379

In his herte was moche sorowe RLH 2382

In herte hym was nothynge well RLH 2396

King Richard þenne become ful off care RLH 6754

But ywis hem was ful wo Horn 54

Þo was Hauelok swiþe wo Hav. 2745

Ase men þat were in hertte wroþe Beves 4152

Þerffore Kyng Richard was ful wroþe RLH 5072

38. *Parenthesis characterizing or naming knight (usually to complete a couplet).*

He was sterne as any lion Beves (Chetham) 772

He semed a man of miche miȝt Guy 874

Gode kniȝtes þai weren bo Guy 902

He semed kniȝt gode & hardi Guy 957

A modi man & ful of pride Guy 1360

Swiþe gode & hardi in fiȝt Guy 1380

A better kniȝt no miȝt man finde Guy 1386

A sterne kniȝt and a Lombard Guy 1408

Þat was þe doukes neve Otoun:
A kniȝt he was of gret miȝt,
Swiþe gode & hardi in fiȝt. Guy 1418–20

A duhtti kniȝt and no couward Guy 1480

Besibbe he was the Duke Segwyn Guy 2132 (Caius)

In fraunce he was borne, Guyes feere Guy 2173 (Caius)

Of Gurmoise þerl sone Aubri Guy 2230

Was non feller into Surrie Guy 2944
In Fraunce y-bore, a knigt wel kinde Guy 2948
Of Almayne a wel gode knigt Guy 2954
Ȝe no haue herd speke of no swiche man Guy 2956
Þat was steward wiþ þemperour.
Knigt he was gode & hardi,
Ac traitour he was, ful of envie. Guy 2960–2
Þe kinges sone of Birrie Guy 2977
A Sarrazin a strong wiþ elle Guy 2984
A strong knigt he was & an Aleman Guy 2986
A Sarrazin of wicked erd Guy 2988
A Sarrazin modi of hert Guy 3502
Þe doukes nevou Otoun Guy 4776
His better wot y non bot te Guy 5112
Þat was þe doukes nevou Loyer Guy 5124
Þat was þe doukes constable Otoun Guy 5223
Of Mohaut he was y-hoten Grimbald Guy 5302
Richard he hete, an onwrast man Guy 5414

39. *Size of host stated.*

Beues in þe sadel lep,
His ost him folwede al to hep
Wiþ baner brigt & scheldes schene,
Þretti þosent and fiftene. Beves 989–92
His ost cam riding him to,
Wel ten þosend oþer mo. Beves 3409–10
Þe nombre was, veraiment,
To and þretti þosent. Beves 4531–2
Two hundred knyghtes assailled Guy,
And him wolde haue sleyn wilfully. Guy 2093–4
Wiþ fif hundred of gode knigtes Guy 5165
& wiþ hem fif hundred knigtes of pris Guy 5172
Wiþ a þousend knigtes & mo Guy 5186
& a þousend knigtes in her compeynie Guy 5193
A þousand knigtes schul wiþ ous gon Guy 5214
Here ffolk were rengyd in þat playn,
Ffoure score þousand, forsoþe to sayn. RLH 4471–2
Almost come ffyue & ffyffty þousinde
Wiþ Saladyn þat comen behynde. RLH 5003–4
Fforþ þey prekkyd, as j ffynde,
Wiþ knygtes fully twenty þousynde. RLH 5027–8
The Sarezynes comen into þe ffeelde
Wiþ sixty þousand & weel moo. RLH 5213–4
Syxty þousand þere were telde
Off gylte spores in þe ffeelde,
Wiþouten ffootmen & putayle,
Þat þer come into batayle. RLH 5387–90
Þer were a þousind knihtes slawen
Bi boþe halue, and mo y-nowe. Hav. 2681–2

40. *How long they fought.*

Þus þe bataile gan leste long
Til þe time of euesong. Beves 797–8
And euer hii were to figt prest
Til þat þe sonne set in þe west. Beves 1015–6
Þus þai leide on in boþe side
Betwene midmorwe & vndertide Beves 1755–6
Þus to gederes þai gonne dinge
Fram prime til vnderne gan to ringe. Beves 4167–8
Þus þat figt leste longe
Til þe time of euesonge. Beves 4451–2
So was bi-twenen hem a fiht
Fro þe morwen ner to þe niht,
So þat þei [stinted] nouht ne blunne,
Til þat to sette bigan þe sunne. Hav. 2668–71
Þe batayle laste tyl it was nygt RLH 5861

41. *Why tell a long story?*

Wharto schuld ich ȝou telle more?
Þe Sarrazins ouercomen wore:
Wele haþ Gij don þat day,
As gode knigt & verray. Guy 3565–8
What schuld y make tales mo? Guy 4140
Wharto schuld y make tale of nougt?
Alle he haþ hem to deþ y-brougt. Guy 4783–4
Whar-to schuld y þer-of lye?
Þat day þai hadde þe maistrie. Guy 5157–8
Wharto schuld ich held long tale,
And michel speke about dualle?
Ac sigge ich-il soþeliche
Wiþ fewe wordes simpeliche,
Gij, & Herhaud, & Tirri þe fre
Wiþ her felawes, þat gode be,
Han ouercomen þe batayle
Atte nende, wiþouten faile:
Alle þe Lombard ben ouer-come,
Oway y-flowen þer be some. Guy 5345–54
What helpeþ hit to make fable? Beves 2147

42. *Summing things up.*

Of al þat oþer, siker apliȝte,
Þat were ensemled in þat fiȝte,
Of Scotlonde and of Al-maine,
Beues and Saber wiþ miȝt and maine
Wiþ deþ is dentes gonne doun driue,
Þat þar ne scapede non aliue.
And þus sire Beues wan þe pris
And vengede him of is enemis. Beves 3441–8

To-ward Mombraunt þei wolde fain,
Ac Saber made hem terne again,
And sire Beues and sire Terry,
And sire Miles and sire Gii
Slouȝ hem doun riȝtes þore,
Þat þer ne scapede lasse ne more. Beves 4247–52

Than miȝt men se fiȝt aginne,
Heuedes cleue vnto þe chinne,
Euerich þat day þat Gij oftoke,
Sone anon his liif forsoke.
Sum he smot opon þe hode,
At þe girdel þe swerd astode;
And sum he smot þurch þe side,
Þat miȝt he neuer go no ride
Was þer non that miȝt astond
Dint þat come of Gyes hond.
So miȝti strokes þer wer ȝiuen,
Þat strong schaftes al to-driuen;
No was þer non in þat ferrede
Þat of his liif him miȝt adrede. Guy 1393–1406

Guy the Almaignes before him wreketh:
Many he taketh, and many he sleeth.
Guy theim driueth, and fast they flee,
As folke that greuously ouercome bee.
Guy (Caius) 2099–2102

On either side they foughte wele
With their launces and swerdes of stele:
They smote of hedes, armes, and honde;
All to-hewen they lye in the sonde.
Guy (Caius) 2167–70

Sone there beganne a strange shoure:
To-geder they smyte knyghtis of valoure.
So many strokes yiuen thou might see
Of the knightis that smote so free;
Bothe with spere and with swerde
They yiue many strokes and harde.
Ther men might see straye many a stede,
And many a knyght shriche and grede.
Whereto shuld y make a tale of nought?
The Almaignes were to deth brought.
Guy (Caius) 2181–90

The Almaignes they goo to assailly,
And with grete strength ouercome bee they.
Tirry to theim was euere meuyng,
And with his swerde gretly harmyng.
Now goo the Almaignes fast fleyng,
And in their fistes their swerdes bering.
The Duke Segwyn ayene come,
And lete theim passe their wey home.
Guy (Caius) 2243–50

Now þai smitte togider comonliche,
& fiȝt þai agin ardiliche.
Þer men miȝt se Gij smite,
& þe Sarrazins heuedes of strike,
& wiþ him Herhaud also:
Boþe þai strengþed hem wele to do.
Þe Sarrazins þai strengþed hem for to sle,
To-hewen, & iuel to bise.
Þe Sarrazins hem ȝeld gret fiȝt,
For strong þai ben, & of gret miȝt. Guy 2965–74

Þe Sarrazins anon gun þai mete,
Mani on þer her liif þai lete,
Mani on þer dyed in aiþer side,
Ac þe Sarrazins wers gan bi-tide.
Sir Gij & alle his feren,
Þe Griffouns þat gode weren,
Han ouer-comen & aqueld:
To-hewen þai leyen in the feld.
Toward her ost þai ben fleinge,
& Gij hem after fast folweinge;
Ar hij þe doun were ouer gon,
Y-slawe hij ben & to-hewen ichon. Guy 3023–34

Þe Sarrazines hij to-heweþ & quelleþ,
Bi þe doun hij gredeþ & ȝelleþ. Guy 3523–4

Gij & Tirri hem folweþ stron[g]liche
& her feren, hardiliche.
Alle þai ben ded oþer ouer-come:
Þer bileued non vnnome.
Gij & Tirri oȝain ben y-gon;
Wiþ hem þai ladde her prisouns ichon.
Whar-to schuld y þer-of lye?
Þat day þai hadde þe maistrie. Guy 5151–8

Longe or the daye began to dawe,
Twenty thousande Gryffons were jslawe.
RLH 2319–20

Vpon here steedes manly þey lepen,
Swerdes & speres to them þey grepen.
Manye man þere slowȝ oþir.
Many a Saresyn loste þere his broþir,
And manye off þe heþene houndes
Wiþ here teeþ gnowȝ þe groundes.
By þe blood vpon þe gras
Men myȝte see where Richard was!
Brayn & blood he schadde jnowȝ,
Many an hors hys guttes drowȝ.
Þere was manye an empty sadyl,
Þat it bewepte þe chyld in the cradyl. RLH 5111–22

43. *None could stand against him (them).*

They went fforth, they had suche mayn,
That none shold stond ham ageyn;
Beves (Chetham) 787–8

Non ne moste oþer abide — Beves 1774

Alle þo þat hii miȝte hitte,
Noman miȝte here strokes sitte. — Beves 3825–6

No strok ne moste oþer abide — Beves 4212

Was þer non that miȝt astond
Dint þat come of Gyes hond. — Guy 1401–2

Nas þer non þat his dint sat — Guy 3498

Was neuere non þat mouhte þaue
Hise dintes, noyþer kniht ne knaue. — Hav. 2696–7

Non þere durste abyde hys dent — RLH 6154

44. *The hero slays many.*

And fifti Sarasins, in þat stonde
Þar wiþ a ȝef hem dedli wonde . . . — Beves 635–6

And þo in a lite stounde
Þe ten forsters wer feld te grounde. — Beves 885–6

And sire Beues, þe cristene kniȝt
Slouȝ ase mani in þat fiȝt
Wiþ Morgelay him self alone,
Ase þai deden euerichone. — Beves 1011–14

Beues and is ost wiþ inne a stounde
Sexti þosent þai felde to grounde — Beves 1017–8

Befyse & hys oost, with owt lesyng,
XV thousande to deþe gan dynge.
Beves (Cbg. U. MS) 1007–8

Beues men myght se in a stound
An hunderid Sarzins ffellid to ground.
Beves (Chetham) 793–4

Þo king Grander was islawe,
Þe seue kniȝtes of heþen lawe
Beues slouȝ þat ilche stounde,
So hit is in Frensch y-founde. — Beves 1779–82

He made after hem so hard chase,
That ten thousand slayn he hase.
Beves (Chetham) 3751–2

And seuen erles he gan doun þrawe,
Sum iwonded and sum y-slawe. — Beves 3819–20

And slowe doun riȝtes mani & fale,
Sixti þousand told in tale; — Beves 3983–4

So þat in a lite stounde
Fiue hondred þai brouȝte te gronde. — Beves 4393–4

Wel mani kniȝtes gij wan þat day,
Of þe maistri he wan þat play. — Guy 889–90

Þurch strengþe of kniȝtes snelle,
Fif hundred of kniȝtes felle. — Guy 5099–5100

Wiþ þat ylke launse selue
King Richard slowȝ kynges twelue. — RLH 5087–8

Off my tale bes nouȝt awundryd:
Þe Ffrenssche says he slowȝ an hundrid.
RLH 5099–5100

Syxe þousand & seuene score
At onys he droff hym before. — RLH 5131–2

Systy þousand þere were slawe
Sarezynys off heþene lawe,
And off Crystene but ten score:
Blyssyd be Jhesu Cryst þerffore! — RLH 5153–6

Many on in a lytel stounde
Wiþ his ax he brouȝt to grounde. — RLH 6149–50

So þat on a litel stund,
Felde he twenti to þe grund. — Hav. 1858–9

But er he were fro Godrich drawen,
Þer were a þousind knihtes slawen
Bi boþe halue, and mo y-nowe,
Þer þe ferdes to-gidere slowe. — Hav. 2680–3

45. *Prisoners are taken.*

He bar hors and man to þe castel — Beves 3440

Wiþ hem þai ladde her prisouns ichon — Guy 5156

He ledde hym to hys pauyloun
Jn despyt of here God Mahoun. — RLH 5127–8

And dide him binde and fetere wel
With gode feteres al of stel;
And to þe quen he sende him. — Hav. 2759–61

He toke them in hostage tho,
There were a thousande prysoners & mo.
RLH 6869–70

46. *The enemy flees.*

Whan they sye that, they fled sone
Beves (Chetham) 3748

To-ward Mombraunt þei wolde fain,
Ac Saber made hem terne again. — Beves 4247–8

He turned his stede & gan to fle — Guy 1511

Guy theim driueth, and fast they flee
As folke that greuously ouerecome bee.
Guy (Caius) 2102–2

Gilemyn with-drowe abacke fleyng
Guy (Caius) 2145

Anon þe Almaundes gin flen — Guy 2247

Toward her ost þai ben fleinge — Guy 3031

Esclander is oway fleinde Guy 3035

Þe oþer oway fleye ich man Guy 5150

Þemperour wiþ þat ginneþ fle Guy 5231

Oway y-flowen þer be some Guy 5354

Wiþ al hys hoost he hym wiþdrowȝ,
And ffleyȝ quyk wiþ hys barounnage. RLH 3168–9

Away þay ryde, & swyþe runne,
Þat nyȝt to ffle & to hyde. RLH 3190–1

Þey ffledde into þe heyȝe tour,
And cryede trewes & parlement. RLH 3230–1

Anon hastyly þey ffledde RLH 4568

Þe Sarazynes fledden, awey gunne ffyken RLH 4777

Þe Sarezynes durste no lengere abyde RLH 5130

He lefte hys pauyloun & hys tente,
And fledde away verramente. RLH 5141–2

Þus þe Sawdon wiþ dolour
Ffledde ffro þe batayle off Arsour. RLH 5151–2

Þe paens þat er were so sturne,
Hi gunne awei vrne. Horn 877–8

47. *None escaped alive.*

Þat þar ne scapede non aliue Beues 3446

Þat þer ne scapede lasse ne more Beues 4252

Þat non off hem ne myȝte ascape RLH 4573

Of alle þe kynges kniȝtes
Ne scapede þer no wiȝte. Horn 885–6

48. *Never before was such a fight.*

Alle þat siȝen hem wiþ siȝt,
Seide, neuer in none fiȝt
So stronge bataile siȝe er þan
Of Sarasin ne of cristene man. Beves 4169–72

Of swiche ne haue ȝe nouȝt y-herd Guy 960

The gretest batayll, j vnderstonde,
That euer was in ony londe. RLH 6883–4

49. *The victors give thanks.*

Þanne al þe folk þat þar was,
Þankede Iesu of þat gras. Beves 2909–10

And þo þe cristen siȝe þat siȝte
Þai þankede god in alle wise,
Þat Beues hadde wonne þe prise. Beves 4240–2

& þankede Iesu, our sauiour,
Þat hadde sent him so gode sokour. Beves 4525–6

Þerl þer-of wel glad he was,
& þonked god of þat gras. Guy 1057–8

And þankyd Jhesu, kyng off glorye,
And Marye his modyr of þat victorie. RLH 4911–2

Gret ioye gan þe Crystene make,
They knelyd & þankyd God off heuene,
And wurschepyd hym & hys names seuene. RLH 5864–6

And þankyd Jhesu, kyng of glorye,
And hys modyr off þat victorye. RLH 7163–4

THE CHALLENGE OF THE PSYCHOPHARMACEUTICALS

NATHAN S. KLINE, M.D.

Director of Research, Rockland State Hospital, Orangeburg, New York

(*Read November 13, 1958*)

THE psychopharmaceuticals present a challenge to (1) those involved with their production and evaluation, (2) to the practicing psychiatrist, (3) to the medical profession at large, and (4) to our social institutions and cultural patterns.

SIGNIFICANCE OF THE PSYCHOPHARMACEUTICALS

In the 175-year history of public mental hospitals in the United States there had always been (until 1956) more patients in the hospitals at the end of any single year than there were at the beginning. This total had then reached the staggering number of 750,00, i.e., three-quarters of a million patients with an estimated cost to our economy of some four billion dollars a year. Based on the projected figures of the last few decades, the anticipated increase in patients during 1956 was somewhere between 10,000 and 12,000 individuals. At an average yearly maintenance cost of $1,250 this would have totaled $12 million to $15 million and if the necessary hospital beds were constructed at an average cost of $12,000 to $15,000 the total would have been closer to $150 million with the prospect that each year this amount would have increased farther. Any reduction in the 10,000 to 12,000 increment of expected patients would have represented a definite improvement. These figures were startlingly revised following the introduction of a new group of drugs. Instead of the total of patients merely being held constant there was, during 1956, an actual *decrease* of over 7,000 patients. This thus represented the equivalent of discharging 17,000 to 19,000 patients. New York State, where progressive leadership made application of the drugs possible at an early date, recorded a 23 per cent increase in the number of discharges. This resulted in a decrease in the mental hospital population of 500 instead of the 2,500 to 3,000-patient anticipated increase. These rates of discharge have now continued into their fourth year so that the burden of proof that this is caused by something other than the pharmaceuticals rests with those who would claim that this is a statistical fluke or attributable to some unknown factor.

WHAT ARE THESE PSYCHOPHARMACEUTICALS AND WHERE DO THEY COME FROM?

Although there are a number of other drugs, I will discuss only three of the major groups of pharmaceuticals used specifically for treatment. *Rauwolfia serpentina* is a common weed in Southeast Asia and other varieties of *Rauwolfia* are found throughout most of the sub-tropical areas of the world. According to Dr. Vakil, whose article on the use of the drug in high blood pressure attracted the interest of Western science, references to its use have been found back as far as 3000 B.C. When claims for its use in the treatment of high blood pressure were verified, we began wondering whether the claims for usefulness in the treatment of "insanity" were also valid ones. About this time the purified alkaloid, reserpine, was made available by Ciba Pharmaceuticals and we therefore tested both a preparation of the whole root supplied by Squibb and the purified alkaloid. Credit for the use of the whole root in the treatment of psychiatric patients belongs to the Indian investigators but our group at Rockland State Hospital was the first to report on the use of the more purified alkaloid. The drug is relatively safe, inexpensive, and has a wide variety of uses. Subsequently, other drugs have been developed which are somewhat more effective for excited hyperactive states but there are many patients for whom reserpine or other *Rauwolfia* derivatives are the drugs of choice.

In an effort to find means of potentiating anesthesia so that smaller doses of possibly injurious anesthetics could be used, Laborit (5)[1] investigated the properties of the synthetic preparation chlorpromazine. He found this to work very well and, since one of the treatments used in psychiatry is prolonged sleep, chlorpromazine was also introduced for treatment of these patients in conjunction with barbiturates and other medications. It was noted that patients whose sleep treatment included chlorpromazine seemed to respond better than patients given ordinary sleep treatment.

[1] Numbers in parentheses indicate references at end of article.

Deniker and Delay (2) then were struck by the possibility that it might be chlorpromazine itself, rather than the sleep treatment, that was so benefiting the patients. They investigated this possibility, found it to be true, and published their results.

The chemical nucleus around which chlorpromazine is built is a phenothiazine and subsequently dozens upon dozens of other phenothiazine derivatives have been manufactured and tested. Some of these are superior in certain respects for certain patients to the original preparation and there are now at least a dozen first-rate drugs in this class which are available for general use. Each of them varies slightly from the other so that each (including the original chlorpromazine) can be used selectively in respect to the needs of the patients.

The third group of drugs has an equally interesting history. Iproniazid (Marsilid) was originally used in 1952 for the treatment of tuberculosis patients and was one of the two "miracle drugs" in this field. However, because of certain side effects, the alternate preparation was generally accepted and the use of iproniazid was virtually discontinued. In the meantime it had been found that iproniazid was an amine oxidase inhibitor and it was used in experiments attempting to clarify the mode of action of reserpine. Since I obviously had a paternal interest in anything connected with reserpine, I was pleased to be invited to see the particular effect described while at the Warner-Chilcott Laboratories. The curious finding was that, when iproniazid was given first and reserpine administered subsequently, instead of the animal becoming quieter he became considerably more agitated. Upon viewing mice so treated, I was impressed by the fact that they appeared to be "super mice." It occurred to me that it would be wonderful if we could get some of our apathetic and depressed patients even to remotely resemble these hyper-alert, hyper-attentive animals. Upon my return I discussed the matter with my colleague, Dr. J. C. Saunders, who pointed out that a year or two before, on theoretical grounds, he had mentioned the possibility of using an amine oxidase inhibitor for treatment. We then placed this on our list of "ideas to be tested," and eventually carried out the experiment a year and a half ago. A researcher always runs an experiment with the hope that his results may be positive. On a hypothesis as remote as this one our expectation of positive results was not high, but as sometimes happens we were blessed by serendipity to the extent that in the first year following our initial report the drug was administered to over 400,000 patients in the United States alone. Since that time a number of other preparations that are also amine oxidase inhibitors are being prepared for marketing, and, as with the phenothiazines, some of them appear to represent improvements over the original compound in respect to usage for particular types of patients. The use of electroshock therapy has markedly declined in many places since, in the large percentage of depressed and apathetic patients in whom the drug works, it is pleasanter, less expensive, and often more effective. As yet we cannot determine by chemical test or any other laboratory means which patients will respond to which medication, nor are results entirely consistent. Until we finally do understand the mode of action, we can only use empirically derived probabilities plus that major fact in the art of medicine—clinical judgment. Whether the mechanism of action is known or not, actual results from the use of the psychopharmaceuticals have amply justified their application.

THE CHALLENGE TO RESEARCH

On the surface the challenge would appear to be: produce compounds which will affect more psychiatric conditions, will be more effective, and have fewer side reactions. To clarify what follows subsequently, I would make division between the major and the minor psychopharmaceuticals. By a major pharmaceutical I mean one that is able to bring about certain therapeutic results which could not be achieved previously in respect to a symptom or disease of a severely incapacitating or debilitating type. By a minor pharmaceutical I refer to one that alleviates an irritating, unpleasant but non-crippling condition, or alternately to a drug which acts only occasionally or to a limited extent in the relief of the more serious syndromes.

There have been only three groups of major psychopharmaceuticals to date: (a) the phenothiazine derivatives starting with chlorpromazine, (*b*) the *Rauwolfia* derivatives starting with reserpine, and (*c*) the psychic energizers starting with iproniazid. The minor psychopharmaceuticals are too numerous to designate in all their detail but meprobamate which affects mild tension and anxiety states illustrates one type. Benactyzine,

which modifies occasionally or to a slight degree a severe condition such as obsessive-compulsive neurosis, illustrates the other type.

Although considerable ingenuity has been demonstrated in evolving new side chains for the phenothiazine nucleus and all of the important alkaloids appear to have been extracted and tested from *Rauwolfia serpentina*, no new or major discovery has been made in the past five years (except iproniazid, which was not new, and the recognition of which as an important psychopharmaceutical occurred accidentally). I would attribute this failure to a lack of *systematic* investigation as to how the major pharmaceuticals do work and to a failure to develop adequate techniques of screening new drugs. Individual investigators in universities, state hospitals, pharmaceutical houses, and other institutions doing research have carried out laborious and at times provocative studies in single areas of drug response, but nowhere has there been a systematic research project dealing with the problem in such a manner that definitive evidence could be obtained as to mode of action.

Parenthetically the point should be made that this is *not* of necessity the responsibility of pharmaceutical houses producing the drug. The minimal expenditure for each drug would be a third to a half a million dollars even if patients were conveniently available. A pharmaceutical house is a commercial enterprise with research justified only if there is good assurance that it will lead to new and useful products. As it is, the pharmaceutical industry spends a higher percentage of its gross for research than any other major industry. If such a company has a successful product, research on its mode of action on such a large scale would be justified only if there were high assurance that a superior product could be evolved. If a company does not have such a product, there is little point in doing such investigations on an ineffective or minor psychopharmaceutical, and to invest this large an amount in the determination of the mode of action of the product of a rival firm would also be difficult to justify. As I proposed recently to a Committee of the United States Congress, I believe that the financing of such basic work should be undertaken by contract with funds supplied by the United States Government, such as is the case with cancer chemotherapy.

In contrast, the development of new screening techniques involves no such mass expenditures and could be undertaken by individual investigators in a large variety of locales. Failure to recognize fully the significance of the need has led to its neglect. Although the chemists are capable of producing thousands of compounds, only a few dozen a year can be put into clinical trial. A really useful screening technique could not only select out the compounds which should be given clinical trial but might also determine which of the available preparations would be most effective in a particular case. In addition, it would unquestionably provide clues as to etiology of mental disease. Both the development of new drugs and assistance in making a diagnosis would result. Even prophylaxis might be possible.

As Leucippus stated a few millennia ago, "Where there is ignorance theories abound." The very plethora of explanations as to the mode of the action of the drugs is indicative of our paucity of knowledge. One of the most widely known theories in respect to reserpine and chlorpromazine is that of Brodie (1) which is set within the general framework of Hess's theory of the balance between the sympathetic and the parasympathetic nervous systems, i.e., ergotropic and trophotropic activity.

In a paper by Ostow and Kline (7) an explanation is proposed in terms of Freudian metapsychology. Psychoenergetics, along with psychodynamics and psychotopology, is one of the three subdivisions of the subject. We would hold for the general thesis that the ataraxics or so-called tranquilizers reduce the amount of available psychic energy and consequently the *lust* or drive behind the disturbances. It is the strength of these drives which causes the setting-up of ego defense mechanisms or the escape from reality. The drugs cause a decrease to the extent that such defenses are no longer so essential and reintegration becomes a possibility. Within the framework of this same theory, the amine oxidase inhibitors increase the absolute amount of psychic energy, making it possible for the individual to dispel or overcome the depression. Drugs such as imipramine offer another possibility within the same frame of reference; psychic energy already available but bound-up in neurotic or psychotic complexes is released to become available for more general use. This will relieve certain types of depression or even specific types of neurotic or psychotic symptomatology. The amine oxidase inhibitors retard the breakdown of a number of normal body chemicals and, of particular in-

terest, that of noradrenalin, which we believe may be essential to maintaining a high level of energy. Let me again emphasize that these are our own postulations and theories which are not necessarily accepted or concurred in by all of our colleagues. Definitely one of the challenges in the field is to explain the mode of action of these valuable compounds.

This lack of knowledge as to the mode of action is the rule rather than the exception in respect to new pharmaceuticals and medicines. After even several millennia of use, we still do not know the mode of action of drugs such as opium or morphine, and, although some 15,000 tons of aspirin are produced yearly in the United States, we again do not understand how even this compound manages to work.

An interesting possibility is that the models, concepts, and constructs used to describe and explain drug action are not suitable. I have elsewhere (3) pointed out that the nature of biological material may be such that the models of the physical sciences have only very limited application. This may even apply to the classification of mental disease (4).

There are a number of stumbling blocks which make it difficult to screen drugs or to determine their mode of action by the use of experimental animals. Probably the foremost of these is that among ordinary experimental animals we see nothing similar to the psychosis or neurosis occurring in humans. As Mettler (6) has pointed out, the only time that animals seem to become neurotic is when they are in contact with humans. Even in those animals in which "experimental neurosis" have been produced, there are striking clinical differences from the neurosis occurring in humans. To test an anti-bacterial or anti-epileptic agent in an animal that does not have an infection or epilepsy presents a difficult problem. One has to look for secondary or tertiary signs, which very often are in the nature of side effects to attempt to estimate whether the drug may or may not have some clinical use. Since traditionally animals are used by pharmaceutical houses for the standardization of certain procedures, the individual who breeds such animals tends to weed out the eccentric and atypical ones. A second process of screening usually occurs when the animal care man receives such specimens, since he will also discard any which show unusual reactions. It is not beyond the realm of possibility that these are the very animals in which the most effective testing could be carried out. If it could be demonstrated that pharmaceuticals known to be effective in humans also restored "off-beat" animals, it might be possible then to administer other substances to determine whether they also possessed similar corrective properties.

CHALLENGE TO THE PSYCHIATRIST

Years ago there was current a story of the uninitiate patient who, after seeing a psychiatrist, expressed chagrin by complaining "What, no medicine?" The advent of medication opens up entirely new possibilities. When psychotherapy was the sole method of office treatment for the vast majority of cases, it was practically obligatory to spend the full fifty-minute hour with each patient, since those who were good psychotherapeutic subjects needed this much time. For those who were not good psychotherapeutic subjects, at least this much time had to be spent to produce any influence whatsoever on the patient's behavior, despite the fact that this did not occur very often. In this manner the custom evolved of spending approximately an hour with each patient. Rumor to the contrary, psychiatry is among the poorer paying specialities, because of the extremely limited number of patients who can be seen during a day at an hour each and the inability of the psychiatrist to delegate or relegate any part of his activity to nurses or technicians, regardless of their competence. This led to the general myth that psychiatrists charged exorbitant fees. This situation was further aggravated since the psychiatrist found it necessary to see the patient usually two to five times a week for a period of a year or longer.

In a significant percentage of patients the one-hour session is no longer actually required but custom dies slowly and, as will be pointed out shortly, another time-honored practice tends unnecessarily to perpetuate it. Now that a more suitable method of treatment is available for those patients who are not good psychotherapeutic subjects, but are amenable to pharmaceutical treatment, there certainly exists no real necessity for spending this much time and seeing the patient so frequently in all cases. Treatment is not entirely dependent upon symbolic communication and the chemicals produce their effects, as elsewhere in medicine, without the psychiatrist necessarily being present. Since there are patients who are poor subjects for psychotherapy, but who do respond to pharmaceuticals, a new treatment situation is created

in respect to office practice. Regardless of how meager or relatively ineffectual it proved to be, providing the patient with "insight" was a major therapeutic objective. Here, too, the full "fifty-minute hour" constituted a bare minimum. The fact that this may no longer be required is disruptive, since it is now possible for the psychiatrist to see three or even four patients an hour if they are being treated primarily with medications.

Many of the patients have also been led to expect the investigation and ventilation of their sexual practices, their bowel habits, and their fantasy life. I have had patients, who despite full symptomatic recovery, were disturbed because they "didn't understand" why they were better. It was necessary to point out that they also did not understand why they recovered from an attack of pneumonia as a result of penicillin treatment. This does not mean that psychotherapy does not serve a useful purpose in these patients, but in many of them depth analysis is not only unnecessary but at times definitely contraindicated. Help is frequently needed in assisting the patient to readjust but as a rule this does not require the time-consuming effort that was necessary in traditional psychiatric practice. The explanation that the self-recuperative powers of the person (as presented by Franz Alexander among others, and more recently by Fritz Freyhan) is encouraged by the use of the drugs seems the most logical and satisfying explanation.

The ability of the average psychiatrist thus to see a considerably larger number of patients also makes possible a marked reduction in fees, putting them in line with those of other specialists. The psychoanalyst or other types of depth therapist would have to retain his present time and fee schedule but would be able to deal exclusively with those patients who might profit by such particular skills. The availability of such treatment in terms of larger numbers of patients who can be seen, as well as larger numbers who could afford therapy, if the psychiatrist were able to see three or four patients an hour, would help to freeze out the nonmedical "fringe-nibblers" whose existence was fostered by the unavailabilty of psychiatric time and the expense involved.

A small but important point of medical ethics is involved in that it would now become possible for the psychiatrist to provide the customary courtesy to his medical colleagues and their families, and not have to insist that psychiatry was so different and so far removed from the rest of medicine that the payment of fees (even for physicians themselves) was necessary for both economic and psychological reasons.

Finally, as will be pointed out in the next section, there are occasions when general practitioners or specialists in other fields can and should use pharmaceutical treatment for psychiatric conditions. Under these circumstances, the psychiatrist will not infrequently be called upon to act as consultant—again, as in any other medical specialty. This means getting out of the habit of regarding a patient as a legacy rather than a referral. After adequate diagnosis, which may take more than one visit as is true in other fields, the patient should be returned to the referring physician with a concise but meaningful summary of his disorder, a diagnosis, and recommendations for treatment which may or may not include specialized psychiatric therapy.

Although the psychiatrist is frequently found bewailing the long list of patients waiting to see him, he is nevertheless reluctant to admit that in certain types of psychiatric conditions the general practitioner can perform just as adequately as the psychiatrist. Objection may also be raised that the communication between the patient and the psychiatrist is "privileged" and should not be disclosed—even to another physician. This constitutes rank snobbery and has entailed justifiable resentment. The patient referred by the psychiatrist to the urologist who reports back on the existence of venereal disease is no more betraying a confidence than the urologist who refers the patient to the psychiatrist for an opinion in respect to the patient's psychosexual status. It would be considered bad form if the urologist were to mention that the venereal disease was contracted from the wife of a mutual friend and the psychiatrist in turn is certainly not called upon to enumerate all the details of a perversion, or to specify the origins of anxiety and tension or of hostility by naming specific individuals or situations when such information is not relevant to the description.

CHALLENGE TO THE MENTAL HOSPITALS

The following are proposed steps to reduce further the mental hospital population through the development of new preparations or by the improved application of those already available:

1. Application of the preparations and techniques now available to suitable untreated patients.

2. Development of new techniques and of combinations of available preparations toward this same end.

3. Investigation of new preparations to see whether they cannot also be used to reduce the chronic hospital population.

4. Utilization of the same techniques to see whether the discharge rate of recently admitted patients cannot be markedly augmented.

5. Application of these same techniques to determine whether many of the patients admitted to mental hospitals could not be treated privately or in clinics to avert the need for such hospitalization.

6. Providing the mental hospital facilities needed so that the patients now institutionalized can be treated with an optimum expectation of successful outcome. This would include not only the necessary buildings, but such essential personnel as psychiatrists, nurses, occupational therapists, etc. Because of the success of the psychopharmaceuticals, *more* rather than less trained personnel in ancillary fields are needed at the present time in order to take full advantage of the possibility of preventing large numbers of such humans from becoming permanent mental hospital residents. The average cost of hospital beds plus an average life expectancy certainly greater than ten years means that the direct cost per patient as things now stand is a minimal $25,000 without counting loss to the community and the humane factors. A temporary rise in expenditures amounting to only a fraction of this is a sound investment economically as well as socially.

CHALLENGE TO THE GENERAL PRACTITIONER AND THE SPECIALIST IN OTHER FIELDS

The use of psychopharmaceuticals should be no different from the use of other drugs. Certain general rules are so familiar as to seem hardly necessary of repetition, yet they have not always been observed in respect to the psychiatric patient. Rule No. 1 would be that a physician should not use either a drug or a technique unless he had a fairly good knowledge of how the medication works and what therapeutic effect might be expected. Rule No. 2 is that the physician should be able to diagnose an illness before prescribing. Rule No. 3 is that, if the practitioner recognizes a case to be beyond his competence, or if the patient fails to respond to what the physician considers appropriate treatment, referral should be made to a specialist in the area of the disability.

Since the terminology and concepts (both neurologic and psychiatric) are so complex, most physicians, anticipating no great need for them in future medical practice, either tended to neglect them in medical school or have long ago forgotten them, as there was little or no occasion to put them to practical use. As a consequence it is more difficult for him to evaluate literature in this field and his diagnostic ability is less sharp than in most other medical areas. Add to this reluctance to refer patients for psychiatric consultation for reasons enumerated in the previous section, and certain difficulties are bound to ensue. A higher than average percentage of patients will be misdiagnosed and consequently not given adequate treatment, the drugs themselves may be used too sparingly or too enthusiastically, with fewer successes and more side reactions than is necessary. Since many psychiatric conditions require prolonged treatment before therapeutic response can be expected, the physician may become needlessly impatient, shifting from one drug to another, and abandoning treatment prematurely. In other cases, treatment may be unnecessarily and untherapeutically prolonged. Fear of side reactions may induce a physician to use a drug of less than the needed potency when adjunctive medications might be available to reduce the side reactions of an adequate preparation, or knowledge and experience with such side reactions could be used to reassure the patient of their temporary necessity. Contrariwise, prolonged usage with disregard of side effects may also lead to therapeutic failure and complications.

The solution is perfectly obvious but difficult to achieve: education of the physician in respect to drug usage and psychiatric diagnosis. This would enable him to determine appropriate treatment and to know when referral to a psychiatrist for diagnosis, consultation, or treatment was indicated. What makes this difficult of achievement is that for many physicians a new or forgotten area of medical knowledge suddenly becomes a necessity. In the midst of a busy practice it is extremely difficult to devote the time and energy necessary, or to make the financial sacrifice required to bring this about. The medical societies have made a valiant effort in this direction but one or two lectures a year at the scientific meetings cannot provide the systematic knowledge needed. The medical journals have included more and

more articles on psychiatric problems, but without the requisite background they may be subject to misinterpretation. Some of the pharmaceutical houses have tried to provide various types of assistance in furnishing the practitioner with the appropriate background. None of these, separately or together, seem likely to provide what is necessary. The one most sensible program advocated is that the federal govenment through the United States Public Health Service, under the direction of the American Psychiatric Association, should provide systematic lectures or refresher courses to be financed by federal funds. A start has been made in this direction but the program needs very definite expansion if it is to achieve its purpose. In the meantime, when in doubt, the diagnostic or consultive abilities of the psychiatrist should be utilized.

It is not the purpose of this article to delineate the types of patients or the recommended treatments which can appropriately be handled by the non-psychiatric medical practitioner. However, it is worth pointing out that a large segment of patients who previously would have had to be either hospitalized or referred to the psychiatrist, can now be treated in the general run of practice. I am specifically referring to the schizophrenic who has been discharged from the hospital on a maintenance dose of a drug and for whom psychotherapy is a waste of time. Adequate medical care of a general nature, plus the ability to recognize whether a significant change is occurring, is the medical management of choice. Another large group of patients of the same type may initially be treated by the psychiatrist, and once maintenance doses of the drugs have been established the patient can be returned to the referring physician with confidence. With the assurance of satisfactory diagnosis and familiarity with treatment routines, the practitioner can retain and appropriately treat a large percentage of the patients who previously had to be referred to state hospitals or the psychiatrist. This is of help to the psychiatrist in turn since his referrals will consist primarily of patients requiring specialized skill and knowledge with a screening out of those cases that present no particular therapeutic challenge or gratification.

THE SOCIAL CHALLENGE AND OBLIGATION

Although they not infrequently tend to be irresponsible, the patients also have an obligation to the physician as well as the other way around. Although a cardiac patient would be most unlikely to double or quadruple his dose of digitalis because "he felt it might help," it is not infrequent that patients on psychopharmaceutical preparations do their own prescribing, either multiplying or dividing their dosage on the basis of what they have heard or read from some non-medical "friend," or else they may prescribe for themselves *ad lib* a variety of additional preparations which "they didn't feel it was even worth mentioning" to the physician. In such cases it should be made clear to the patient that if he, the patient, is going to do the prescribing, he will also have to assume the responsibility, and that the legitimate physician will be forced to dissociate himself from the case if he is not allowed to retain his established and necessary prerogative of determining what medication shall be used and in what manner.

Cautious experimentation should be begun in the development of treatment facilities more appropriate to drug usage than the isolated mental hospital built for long-term occupancy. Acceptance of psychiatric patients into general hospitals, the establishment of "half-way houses," the adaptation of various social agencies to the patient and family problems, and solutions, created by the use of the drugs is already overdue. The details of such planning is beyond the sphere of my competence, but it is obvious that there exists no widespread recognition that a problem even exists.

The handful of psychiatrists in the world are helplessly inadequate to cope with the stupendous problem if they utilize only traditional methods. However, the introduction of the pharmaceuticals has opened a new avenue approach in such countries as Haiti, with one psychiatrist (and he is also Minister of Foreign Affairs) for 4 million population or a country like India, with approximately forty medically trained psychiatrists for a population of some 400 million. With the support of three of the major pharmaceutical houses, Hoffmann-LaRoche, Schering, and Wyeth under the auspices of Medico, we are completing the establishment of the Haitian Psychiatric Institute, which will be the first such clinic in the world to build its program around the use of pharmaceuticals as the major method of treatment. We are training part of the personnel for this project at Rockland State Hospital at the present time. If this pilot project proves suc-

cessful, it will open the way to the establishment of similar inexpensive units throughout the world.

It is Standard Operating Procedure that technical advances outstrip social adaptation, but failure to attempt a solution by ignoring the problem constitutes social irresponsibility. Perhaps familiarity with the research, the psychiatric, and the general medical area gives me confidence to predict that within a reasonable period of time the appropriate medical adjustments will occur. It would be encouraging to know that social adaptation would proceed apace. This should be the subject of major conferences by groups such as the Council of State Governments and possibly a national conference by the federal government under the auspices of the American Medical Association.

REFERENCES

1. Brodie, B. B., and Shore, P. A. 1959. Mechanism of action of psychotropic drugs. *Psychopharmacology frontiers.* Little, Brown & Co.
2. Delay, J., Deniker, P., and Harl, J. M. 1952. Utilisation en therapeutique psychiatrique d'une phenothiazine d'une action centrale elective (4560 RP). *Anna Med. Psychol.*, 112.
3. Kline, N. S. 1953. Samples and controls in psychiatric research. *Psychiatric Quart.* **27**: 474–495.
4. Kline, N. S., and Gerard, D. L. 1953. Taxonomy of mental disease. *Jour. Gen. Psychol.* **49**: 201–207.
5. Laborit, H., Jaulmes, C., and Benitte, A. 1952. A few experimental aspects of artificial hibernation. *Anesth. et Analog.* **9**: 232–239.
6. Mettler, F. A. Personal communication.
7. Ostow, M., and Kline, N. S. 1959. The psychic action of reserpine and chlorpromazine. *Psychopharmacology frontiers.* Little, Brown & Co.

THE COMMON LAW AND CODIFICATION IN JACKSONIAN AMERICA

PERRY MILLER

Professor of American Literature, Harvard University

(*Read November 14, 1958*)

THE new nation, the United States, was aware in the 1790's that many leaders of the Revolution and authors of the Constitution had been trained to the law. Nevertheless, the mass of the people survived these adventures with a deep-seated distrust of the law, even with an active hostility. Their suspicions were inheritances from colonial experience, which may be summarized as the conviction of a simple, hard-working, and generally pious commonalty that lawyers, when not composing the Declaration of Independence, but when trying cases in court, were clever sophisticates who made the worse appear the better cause and repeatedly hoodwinked virtuous citizens.

The *Autobiography* (so-called) of Davy Crockett summarized in the age of Andrew Jackson the popular wisdom. He recounted his service as justice of the peace when he dwelt "only a little distance in the purchase, and no order had been established there." And he was successful. None of his acts were ever repealed by a higher court, because, he averred, "I gave my decisions on the principles of common justice and honesty between man and man, and relied on natural born sense, and not on law learning to guide me." He boasted, and a nation applauded, "I had never read a page in a law book in all my life."

But even had a Davy Crockett of about 1790 wanted to read a page, he could hardly have found a "law book" that would instruct him in American terms. He could, of course, easily have picked up Blackstone's *Commentaries*. The nascent profession took as a compliment Burke's warning during the pre-Revolutionary debates that more copies of Blackstone had been vended in the colonies than in Great Britain. Yet all were aware, after the victory, that a great part of Blackstone's four volumes applied only to a feudal, aristocratic, royal order, that they were primarily concerned with problems of land and inheritance; whereas in this wilderness we had too much land to bother with litigation about it, and the mass of people had their inheritance in the shape of political freedom. Hence there was reason for ordinary men to sneer at booklearning and to object against the profession that it constituted a conspiracy against self-evident principles of common justice.

It is an over-simplification of American history to assume, as many have assumed, that opposition to the mystery of the law came only from the Jeffersonian and Jacksonian democracy, and that it was countered only by Federalist and Whig intellectuals. At the beginning, I think it more fair to say, the masses were vaguely aware that nobody quite knew what was or was not law in America, and so concluded that the lawyers were trying a trick of legerdemain, of creating something out of nothing, which something would then enable the profession to prey on the community. Jurists assume, said P. W. Grayson as late as 1830 in a book eloquently entitled *Vice Unmasked*, that it is the function of the people "to be content that all its precious mysteries, its wondrous depths and beautiful harmony, would be unfolded; and to them [i.e., lawyers] alone, the gifted receptacles and conductors of the light of justice, it were becoming in us to rejoice that we were privileged to receive it." Hence jurists bewilder us with "despicable trash and superstitious jargon."

As lawyers looked back on this period from the eminence they attained in the course of three decades, they wondered how they had managed to survive. James Kent remembered that when he went on the bench in 1798 there were still no reports or state precedents; all the time he was Chief Justice, no counsel could quote opinions of his predecessors. The history of how the profession lifted itself out of this bog has been frequently told. At first their main device was securing systematic publication of decisions, of which Johnson's *Reports* for the State of New York and Wheaton's and Peter's volumes for the Supreme Court became the models. Indeed, by 1820 the cry was already loud from students that too many reports existed for the mind of any single man to encompass. Secondly, the law schools, especially Tapping Reeve's at Litchfield

and after 1817 the Harvard Law School, helped bring order out of chaos. And thirdly, and in the period most effectively, came the systematic textbooks produced by David Hoffman in Baltimore and Philadelphia, Kent's *Commentaries* (published between 1826 and 1830, with continued revisions) and Joseph Story's crushing array of commentaries on such subjects as bailments, equity, constitutional law. As early as 1817 Story could note that there were lawyers still living who remembered the degraded state of the calling a few years earlier, though now it had become fashionable. "And this circumstance," he continued, "combining with the nature of our political institutions, . . . has contributed in a very high degree to that great increase of the bar, and that ascendancy in society, which distinguish the profession, in this, more than in any other country."

For men like Hoffman, Kent, and Story, the secret of their achievement was that they had so presented the Common Law of England as to demonstrate its adaptability to American conditions, and in their opinions and commentaries so naturalized it as to overwhelm the unlettered resistance of the vulgar. Story wrote an English judge in 1840, asking him, rhetorically, whether England could point to any nobler triumph "than the proud fact that her Common Law exerts a universal sway over this country, by the free suffrage of all its citizens?"

In this spectacular campaign, the lawyers insisted to the American people that while the Common Law had grown up haphazardly, through the accidents of history, still as it now existed it displayed the perfection of rational symmetry. Possibly it had not been so in the days of Coke, but Blackstone had shown how to organize it as a science. Blackstone, proclaimed Josiah Quincy, waved a magic wand over the illimitable ocean and from its depths summoned up

> a well-proportioned, well-cemented fabric, pleasing to the sight, satisfactory to the taste, approved by the judgment, its architectural principles just, its parts orderly and harmonious, in which justice was found consorting with reason, and controversy guided by the spirit of truth, and not the spirit of victory.

The comprehensive treatises of Hoffman, Kent, and Story were, in the eyes of most of their colleagues, ample demonstration that this conception of the Common Law had successfully survived transition to America. It is not because it is English that we should receive it, said Joseph Hopkinson of Philadelphia, "but because it is the law of *reason* and *justice*."

Yet at the very moment the triumph seemed entirely won, the lawyers were confronted with another home-grown idea which to most of them threatened the utter destruction of their mighty fabric. They were challenged by a horrible Caliban who answered to the alias "codification." Here was a more insidious foe than the earlier and relatively primitive antipathy to law of any kind and to the Common Law in particular, though some of those who had been vociferous against the Common Law reformed their lines under this new standard. It was more insidious because it took the lawyers at their word; it said in effect, "All right, let us grant that you have Americanized the Common Law and that you present it to us as a rational science; why not then reduce the science to explicit and systematic statement, as does, for example, a textbook of physics or chemistry?" The greatest luminaries of the profession were appalled by the very notion, and instinctively organized themselves into a solid phalanx to annihilate the proposition.

There can be no doubt that much of the momentum in the cry for codification came from partisans of Andrew Jackson, and assuredly the most powerful counterattacks were led by Whigs. Some historians see in the entire dispute only a series of political maneuvers. Party passion is obviously an important aspect of the controversy, and is illuminated, to pick only one passage out of thousands, by a letter from a Massachusetts Democrat to Rufus Choate in 1831:

> Whenever a system of plain and equal law has been proposed, the brotherhood of the bar have always with one consent, in the legislature, in the newspapers and every where, represented it as chimerical, as the scheme of levellers; that the English common law is the height of perfection, and that the united wisdom of the world could not produce its equal.

However, Robinson continued, suppose we use our own reason, refuse to be dictated to by interested men; we shall certainly find "that a plain, equal, humane, and concise code of civil and criminal laws, could easily be made and compiled by disinterested legislators." As the dispute grew in vehemence in Massachusetts, a Democratic newspaper issued a call for the people to arise, to overturn the usurpations of those "who have fastened upon us a government of a *privileged order* of society, who rule us as they please, under

an unconstitutional common law, and a chancery law jurisdiction established in violation of the most explicit and solemn guarantees."

Yet there is evidence that these were good Democratic lawyers who were as frightened as was Kent by the danger, and conversely that there existed, at least in the early years of the argument, Whigs who could discuss the issue without suffering strokes of apoplexy. There were several factors which cut across merely political alignments. For one thing, the concept of codification became confused in many minds with the image of Jeremy Bentham, who had criticized the great Blackstone and had in 1817 issued his book called *Codification*. Then he had the effrontery to offer President Monroe his services in codifying all American law, on the bright notion that in a young country there would be less antique lumber to clear away and that a republic would see the point of a rational order where an aristocracy could not. Thousands of patriot hearts indignantly resented Bentham's attempt to meddle. Besides, by the year of his death, 1832, Jeremy Bentham's name had come to stand for everything that in America was held mean, vile, niggling. Reasons for this opinion can be found in many areas of the American society, but it is enough to say that in a nation so highly dedicated to the utilities of improvement and expansion, such was the dominant religious and sentimental mentality that the Utilitarianism of Bentham and James Mill was regarded as a philosophy fit only for pigs. At the head of the insurgency, said *The United States Law Journal* in 1826, stands "that crazy philosopher and arch codifier Jeremy Bentham; a man, in our judgment, so thoroughly mad that we really think it would be in justice to shut him up and keep him on bread and water until he should be brought, by depletion, to some tolerable measure of common sense." Those who agitated for codes took care to dissociate themselves from any taint of Benthamism.

The most slashing champion of codification was William Sampson, who horrified the members of the New-York Historical Society in December, 1823, by asserting that we could settle all doubts on every topic by positive enactment and thus get rid of "those borrowed, ill-penned statutes, of which every word gives rise to endless commentaries." A more thoughtful and so more effective advocate was Thomas Grimke of Charleston, South Carolina, who in that state had some practical influence. It was he who first, or at least most distinctly, voiced the inspiration which kept the codifiers active. The necessity and obligation of reform lies heavily upon young states like the American, which are "rushing forward in the career of improvement, with a velocity unparalleled in the History of Man." Such states, which must be ever mindful of the claims of posterity, will violate their own best interests "if they do not purify, reform and remodel their antique, motley System, the growth of centuries under a Monarchical Government." As a matter of fact, Grimke scored heavily by pointing out that our written constitutions, state and federal, are really codes: "They are the light of the World, the Hope of the Nations, and the rich Inheritance of all Posterity." They sifted and selected, choosing the good, rejecting whatever does not conform to the spirit of this age. "Why then, may we justly enquire, should not a Similar Reform be accomplished in the Civil Department of the Common Law?"

For a short time in the 1820's the question could be discussed with equanimity. *The North American Review*, for instance, ran in 1824 a judicious survey of both sides by Henry Dwight Sedgwick. Joseph Story never quite lost his temper about it, and for Francis Lieber's *Encyclopaedia Americana* in 1836 he wrote an article which honestly respected his opponents. Indeed, when the agitation in that year became so intense that Governor Everett could not ignore it, he appointed Story to a commission which, though advising against wholesale codification, did point out areas, such as criminal law, property and personal rights, contracts, where it might be not only feasible but desirable. Yet for the most part, the knights of the Common Law lashed the codifiers with verbal onslaughts that may accurately be described as hysterical.

There is not space or time to recapitulate here all their oratory. I find it often amusing, however, as it exhibits how these professed devotees of logic and reason, these pundits who constantly were lecturing the common people about emotional instability, could themselves lose all sense of restraint and accuse their enemies of crimes darker than murder or treason. We should not be surprised, I suppose, to discover that Daniel Webster was one of the earliest to take up the cudgels, writing in 1818, "It yet appears to us to be among the idlest and weakest theories of the age, that it is possible to provide, beforehand, by positive enactment, and in such manner as to

avoid doubts and ambiguities, for all questions to which the variety of human concerns gives rise." This thesis would continue to be the principal argument of the anti-codifiers; as time went on, the tone became more shrill. Men like Sampson and Grimke, said a writer in *The United States Law Journal* of 1825, pester us with "their long and wearisome dissertation upon the civil law and the code of Justinian—the code Napoleon —and the Prussian code—and the whole army of codes and commentaries of which they talk so learnedly": they succeed only in showing that in this, "as it is in almost everything else between the English and the American on the one hand, and the French on the other," the French "wear the ruffle but we wear the shirt." Beginning in 1827 *The North American Review* ran a yearly blast against codification; some of these lapsed so far into a violence of vituperation as to seem quite out of place among its otherwise somberly rational pages. The proposal that state legislatures should enact such chimerical codes excited insulting remarks about the intellectual incompetence of legislators. When reformers said that judges legislated by interpreting the Common Law capriciously, their defenders indignantly recited the virtues of judicial restraint. When no other appeal seemed to work, such a book as John Anthon's *The Law Student* (a textbook widely used for many years after its publication in 1850) categorically stated that the destruction of the Common Law "by a set of stiff apothegms, dignified by the name of a code, would be a national calamity."

It is easy for us, looking back on the arena, to surmise that the lawyers, while celebrating the flexibility, the adaptability—what they like to call the "malleability"—of the Common Law, and proving that only highly trained professionals could manage to use it, were defending their prestige in an aggressively democratic society. Their scorn of the ability of legislatures to construct codes that theoretically would cover in advance all conceivable exigencies shows itself often to be their contempt for the representatives of the people. Yet, while their resistance perhaps had certain anti-democratic overtones, there was another plea they persistently advanced in which their sincerity shines more brightly. By their lights, they were as patriotic, as thoroughly American, as their critics. They did not consider themselves obsequious to English precedent. On the contrary, they insisted, after the Revolution they had indeed been obliged to start with a bare minimum of native lore and so had borrowed from England; yet in an amazingly short time they had so domesticated the Common Law that it had become Americanized. The forms of our jurisprudence, said *The North American Review* in 1828, are purely and exclusively our own, and our Supreme Courts, state and national, "have neither precedent nor parallel in any country or age." So, on the verge of Jackson's election, this defender could heap praise upon the courts of the Republic in language which, I feel confident, may arouse an echo of sympathy today:

> We are now witnessing their powerful influence over our political character, in the protection which they give to individual rights against the encroachments of the legislative and executive combined. . . . In the vast machinery of our national affairs, they are, as it were, the regulators. All our great public functionaries, even Congress itself, as well as the legislatures of the several states, they hold in salutary check. Here they are supreme, and exhibit before us continually, in actual practice, as well as in beautiful theory, the absolute sovereignty of the law.

This in substance was a vision of America as dear to these lawyers as any chanted by a Walt Whitman or given form in the "sublime" landscapes of the Hudson River painters.

But what, we still have to ask, were the motives of the codifiers? Were they merely partisans of Andrew Jackson trying through the imposition of codes to strike down, or at least to curtail, the power of Whig judges? Doubtless such an explanation is true in part; it is noticeable that the chief advocates were Democrats. But to take this as full explanation is to settle for a shallow comprehension of the American mind in that era. They were also sincerely convinced, as Campbell P. White said in moving for codification at the New York Constitutional Convention of 1846, that it was incumbent upon America to achieve "a simplification and reduction of the antiquated, artificial and ponderous forms of legal and equitable proceedings—and ultimately for the enactment and codification of the vast mass of unwritten law and equity—in order that the people may know the legal and equitable rules by which they must be governed." Of course, they said, they never intended that the codes should be formulated on the floor of the legislatures. There were plenty of able lawyers, some of whom might know as much about the Common Law as Joseph Story,

who would prepare them with care. The need was urgent: opponents gloss over the discordance and incoherence, the difficulty, delay, and expense of the prevailing methods, yet all the time proclaimed the law a learned science which only a few experts understand. If it be really what they say it is, said Timothy Walker in a Phi Beta Kappa address of 1850 at, of all places, Harvard University, then surely "these rules and principles can be collected together, arranged in systematic order, expressed in precise language, and clothed with legislative sanction, which not one tithe of our law ever yet had." The conception itself had such an inherent beauty, what some went so far as to call a sublimity, that it made dedicated warriors of its converts.

The most persistent of these was David Dudley Field of New York. For him codification was something far beyond a utilitarian simplification: it was a religion. He agitated ceaselessly and was attacked furiously. He prevailed on the legislature to appoint a commission in 1847, out of which emerged his reports on Civil and Criminal Procedure; a second commission appointed in 1857 submitted several reports which were acclaimed as "the Field Codes" and proved more popular in the West, especially in California, than in New York. In the latter state, in fact, the real battle over the Field Codes did not come until the 1880's, when the situation was entirely different and very little of Field's original inspiration was any longer exerting an effect on either side.

Hence it is not with the history of American jurisprudence that I am here concerned. Dean Pound's judgment is that Kent, Story, and their allies deserve the gratitude of the nation because by their resistance they held the movement at bay until by the time of Field the Common Law had been thoroughly digested into American jurisprudence. They saved us from the horrors of premature codification, which might well have become a shirt of Nessus. And it might be said—I speak subject to correction by the jurists here present—that though the codifiers won only a limited success at best, the issues they raised are still unsettled business.

Wherefore I come back to what for me, as a student of American culture, is the pressing question: what was it that lured the codifiers? Here Field, as the most prominent and persistent, as well as most articulate, is most revealing. His numerous pieces of propaganda ring changes on the standard arguments that Common Law procedures are cumbersome, expensive, ultra-sophisticated, and he never tired of denouncing the folly of separate actions for the same suit in both law and equity. Yet in many of his more heated moments he shows effectively that what he sought was a realization of the American personality. He, a New England born Puritan, had become a New York Democrat; he had visions of an America realizing its destiny which were as intense as those of his fellow and younger Democrat in New York, Walt Whitman. There are times when his characterization of the Common Law seems merely an analogue in legalistic prose to the vaunts of literary independence which Melville was putting into his romances and Whitman would make into a barbaric yawp. Here, for example, is Field on the Common Law:

> . . . a most artificial system of procedure, conceived in the midnight of the dark ages, established in those scholastic times when chancellors were ecclesiastics and logic was taught by monks, and perfected in a later and more venal period, with a view to the multiplication of offices and the increase of fees, was imposed upon the banks of the Hudson and the quiet valley of the Mohawk.

Herman Melville, in the ecstasies of his patriotism, prophesied that America, once it transcended its provincial cringing to England, would become the John Paul Jones of nations, "civilized in externals, but a savage at heart." Whitman encompassed all America into *Leaves of Grass*, though this meant pages and pages of inventory of things American, and at the end looked to a future beyond America in "Passage to India." In the very year in which Whitman's book was first published, in 1855, David Dudley Field already foresaw the same glory:

> . . . a CODE AMERICAN, not insular but continental, as simple as so vast a work can be made, free in its spirit, catholic in its principles; and that work will go with our ships, our travelers and our armies; it will march with the language, it will move with every emigration, and make itself a home in the farthest portion of our own continent, in the vast Australian lands, and in the islands of the southern and western seas.

In the history of American law, conceived simply as a technique, the war over codification in the early nineteenth century may figure as a curious waste of energy. The graduate of a

modern law school, disciplined in the case method, perhaps would find incomprehensible the very terms in which it was debated. Yet I suggest that a basic problem for American culture was weirdly raised in the guise of legal philosophy. The Common Law practitioners considered themselves as thoroughly American as anybody, but it was consistent with their patriotism that America acknowledge her relation to the mother country, long after she had attained political independence, by gradually and carefully adapting to American developments the ancient skills. Only thus did they believe America could become a civilization, by remaining within the context of an international culture, certainly not by seeking for identity within its isolated and undisciplined bosom. Only thus could it move from the rough and tumble of the frontier to a society of law and order.

To a greater or less degree, in David Dudley Field preeminently, the codifiers wanted to complete the political Revolution by enacting a still more drastic break in the law. They resented the idea that a foreign system should be imposed upon the beautiful landscape of the Hudson and the Mohawk. They were resolved that the great prairies and the western mountains should not be so defiled. Among the lawyers as among the writers something was at stake which transcended the particular forms or the language: both were seeking within their parallel competences to ascertain the meaning of America.

BERNARD SHAW, SIDNEY WEBB, AND THE THEORY OF FABIAN SOCIALISM

GEORGE J. STIGLER

Professor of American Institutions, University of Chicago

(*Read November 14, 1958*)

THE transition of public policy in England from one of relatively pure *laissez faire* to one of collectivism began in the first half of the nineteenth century, and presumably has not yet reached its end. The shift in public opinion and in effective electoral power which lies behind the shift of policy, therefore, cannot have been initiated by the Fabian socialists, who began their labors in 1884. Yet they are commonly credited with a leading role in persuading the intellectual classes of England of the undesirability of organizing economic life on the basis of private enterprise.

The two leading theoreticians of Fabian socialism were Shaw and Webb—indeed one is inclined to say that they were the only theoreticians in the first decades of the Society. Webb's labors in both economic scholarship and politics are well known; at least in this country there is some tendency to underestimate Shaw's part, simply because his other activities eventually overshadowed as well as displaced his Fabian period. Shaw felt differently:

> Now gentlemen, I am really a political economist. I have studied the thing. I understand Ricardo's law of rent; and Jevons' law of value. I can tell you what in its essence sound economy means for any nation.[1]

The propriety of this claim is better judged at a later point.

I propose to discuss only two aspects of the Fabian movement. The first aspect is the early work in economic theory by Shaw which commands our interest because of its importance as well as its authorship. The second aspect is the precise nature of the theoretical critique of capitalism to which Shaw and Webb devoted their immense talents and energies in the first two decades of the Fabian society.

1. THE EARLY SHAW

Bernard Shaw was first persuaded of the need for radical economic reform, he tells us, when—in 1882—he accidentally drifted into a London hall and heard one of Henry George's influential lectures for a tax on the rent of land.[2] The study of *Progress and Poverty* soon led Shaw toward socialism. Shaw's incomplete novel, *An Unsocial Socialist*, revealed that within a year after hearing George the conversion to socialism was complete. Shaw's hero, a Sydney Trefusis, abandoned his wife and his inheritance—both of admirable dimensions—to devote his days to long speeches on the iniquities of capitalism and to inciting the rural proletariat to trespass. That this prince of prigs did not choke off socialism in England is itself one indication that the novel was not widely read.

The criticisms of Henry George by English Marxists drove Shaw to the French edition of Volume I of *Das Kapital*. He was captivated without being persuaded of the validity of all its economic theory. These doubts spilled into print in a letter to a weekly, *Justice*, entitled "Who is the Thief?"[3] It is a tribute to Shaw's penetration that he had found for himself a crucial flaw in Marx's labor theory of value.

Marx's central argument was that the capitalists, by their control over capital equipment and the means of subsistence, forced a worker who added 10 shillings of value to 10 shillings of material, to work for only 3 shillings (his assumed subsistence requirement), yielding up 7 shillings of surplus value.

> But mark what must ensue. Some rival capitalist, trading in tables on the same principle, will content himself with six shillings profit for the sake of at-

[1] Shaw, G. B., The case for equality, 11, an address delivered on the first of May, 1913. First ed., London, 1913, National Liberal Club Political and Economic Circle, *Transactions*, pt. 85; reprinted in *The socialism of Shaw*, ed. with introduction by J. Fuchs, 58, New York, Vanguard, 1926.

[2] Henderson, Archibald, *George Bernard Shaw: man of the century*, 215, New York, Appleton-Century-Crofts, 1956.

[3] Signed Larking, G. B. S., Reprinted by R. W. Ellis in *Bernard Shaw and Karl Marx; a symposium: 1884–1889*, New York, printed by Random House for R. W. Ellis: Georgian Press, 1930.

tracting custom. He will sell the table for nineteen shillings; that is, he will allow the purchaser one shilling out of his profit as a bribe to secure his custom. The first capitalist will thus be compelled to lower his price to nineteen shillings also, and presently the competition of brisk young traders, believing in small profits and quick returns, will bring the price of tables down to thirteen shillings and sixpence.[4]

But if the worker is being robbed of 7 shillings, then the purchaser is committing thirteen-fourteenths of the theft—every English consumer is the thief. The criticism received no reply.

The assumptions of competition and of surplus value are indeed incompatible, and even today I would like to amend Shaw's argument in only two respects. The competition of capitalists would also take place in the labor market, and force wages up. And, secondly, the customer-thieves are, of course, chiefly the workmen and their families.

The distrust of Marx's value theory was strengthened by an attack made by Philip H. Wicksteed.[5] Using the recently developed marginal utility theory, Wicksteed showed that Marx's theory was illogical. It was illogical because Marx insisted that only socially necessary labor governs values, which introduced surreptitiously the very quality of utility which he had denied as a universal attribute of commodities. The theory was incomplete because it could not cope with the value of commodities which were not freely reproducible (such as old masters) or were monopolized. Wicksteed's own radical leanings (which diminished subsequently) were perhaps revealed by the fact that he did not comment upon the crucial flaw in Marx's theory—the denial of productivity to resources other than labor.

Shaw took upon himself the writing of a good-natured rejoinder to Wicksteed.[6] Ignoring Marx, he rashly attacked the marginal utility theory, a task for which he was unprepared. The attack centered upon the fact that the amount of utility obtained from an increment of a commodity fluctuates widely over time for one man, and varies widely among men, without any corresponding variation in the value of the commodity. The criticism failed to distinguish positions of equilibrium from those of disequilibrium, and Wicksteed had no trouble in disposing of it.[7]

The debate now shifted to a small discussion group, the Hampstead Historic Club, where Shaw and Webb, as well as other critics of capitalism, were joined by two economists, Wicksteed and Edgeworth. For two years Shaw was subjected to training in economic theory by two of the world's leading theoreticians—although one may conjecture that with students such as he the professors learned a fair amount about debating. He emerged a complete convert to Jevons and Ricardo—an odd set of intellectual parents considering Jevons' vast dislike for Ricardo's theory.

The conversion was announced by three notices on *Das Kapital* in *The National Reformer* (1887).[8] They express a deep appreciation of the powerful influence of Marx's denunciation of the injustice of capitalism and of his presentation of a law of historical evolution which gives little more time to capitalistic society. The crucial weakness of the Marxian theory of value proves to be the same point that Wicksteed made, that only relative utilities can account for the observed phenomena of value, and this is tacitly recognized when Marx refers to socially necessary labor.[9]

It will suffice merely to mention Shaw's remaining work, because we shall consider its theoretical content below. The essays on the economic theory of and the transition to socialism in *Fabian Essays* (1889) were his. Thereafter, his economic writings took the form chiefly of Fabian pamphlets,[10] although he also wrote a book on municipal trading.[11] The much later *Intel-*

[4] *Ibid.*, 5–6.

[5] Wicksteed was adding to his careers in the Unitarian ministry and literature—he was the translator of Dante—that of economist.

[6] Shaw, G. B. *The Jevonian criticism of Marx*, 1885, reprinted in Ellis, *op. cit.*

[7] Wicksteed, "A Rejoinder," reprinted in Ellis, *op. cit.* Shaw subsequently wrote that his reply proved nothing but his incompetence; Ellis, 138 n.

[8] Shaw, reprinted in Ellis, *op. cit.* The essays contain a remarkable vilification of H. M. Hyndman, the Colonel Blimp of English Marxism.

[9] Apropos of the "transformation" problem of which there was promised a solution in the third volume of *Das Kapital*, Shaw observes that "scientific socialism" means cashing a promissory note of Mr. Engels, dated "London, an Marx' Geburtstag, 5 mai, 1885" (Ellis, *op. cit.*, 108).

[10] They are identified in E. R. Pease, *History of the Fabian Society*, Appendix IV, new and revised edition, New York, International Publishers, 1926; some are reprinted in *The socialism of Shaw*.

[11] *The commonsense of municipal trading*, Fabian Socialist Series No. 5, London, A. C. Fifield, 1908. Although the volume is of more interest to political scientists than to economists, it has some remarkably bold analyses resting on the differences between social and private costs. Shaw places extraordinary weight upon the ability of cities to borrow at low interest rates.

ligent Woman's Guide to Socialism contains nothing new on our main subject.

Sydney Webb's prodigious literary output seldom lacked relevance to enonomic theory, but only a few early items are germane to our inquiry.[12] They will be considered in the Fabian critique of the basic logic of a capitalistic (private enterprise) system, to which we now turn.

2. THE ECONOMIC THEORY OF FABIANISM

The main Fabian indictment of private enterprise rested squarely upon the classical theory of the rent of land.

> On Socialism the analysis of the economic action of Individualism bears as a discovery, in the private appropriation of land, of the source of those unjust privileges against which Socialism is aimed. It is practically a demonstration that public property in land is the basic economic condition of Socialism.[13]

The crux of this theory is the assumption that, in a settled country, the effective supply of land is fixed. No degree of expansion of the demand for the products of land would call forth an additional acre of land—putting aside normally trifling amounts due to drainage, irrigation, and the like. The aggregate income of the landowners will therefore wax with the growth of population and wealth.

This doctrine had provided the main theoretical support for free trade in grain—the repeal of the Corn Laws in effect greatly increased the supply of land available to meet the demands of British consumers.[14] This was the main goal of policy of the Ricardian school, and one might have expected that once the goal was achieved, the landowner and the rent theory would have receded from the stage of public controversy.

For a time after the repeal of the Corn Laws the criticisms of the private ownership of land abated, but they revived in the latter decades of the century. John Stuart Mill threw his immense prestige behind the Land Tenure Reform Association,[15] and he was succeeded by A. R. Wallace, the co-discoverer of the Darwinian theory. The importance of the absentee rents in the Irish agitation contributed to this revival. The interest in land taxation or nationalization became almost universal after Henry George's triumphal lecture tours beginning in 1882.[16]

Shaw restated the Ricardian rent theory without appreciable modification so far as land is concerned, embellishing it only with his handsome prose. Shaw's piece on the economics of socialism in the Fabian essays was a straightforward reproduction, on a nonrigorous level, and his tract on *The Impossibility of Anarchism*, where anarchism virtually means competitive private enterprise, turns on the fact that it is intrinsic to economic life that some land is more valuable than others.[17] Webb's acceptance was as complete.[18]

How could one seriously make a heavy indictment of private property in land when free trade had made wheat—and land—cheap? Late Victorian England was an odd place to offer heavy criticism of feudalism. The answer is a complex one. In the first place, the facts were largely ignored and, when recognized, misinterpreted. The famous Fabian tract, *Facts for Socialists*, estimates rents to be £200,000,000 or one-sixth of the nation's income in the mid-eighties. Aside from a myriad of minor errors, which the incredibly complex English income tax law invited,[19] the Fabians blithely included the value of all buildings in that of land. The true rent of land (urban as well as rural) surely did not exceed £60 million, or a mere 5 per cent of income. Moreover, agricultural rents had been declining in absolute amount since the 1870's, and as a share of national income they had

[12] There seems to be no doubt that Beatrice did not contribute to the work on economic theory. Perhaps one reason was her haste: "I went straight to the club and read right through Marshall's six hundred pages—got up, staggering under it. It is a great book, nothing new—showing the way, not following it." (Diary, July 27, 1890, quoted in Margaret Cole, *Beatrice Webb*, New York, Harcourt, Brace, 1946.)

[13] Shaw, in *Fabian essays*, 24, Jubilee edition, London, George Allen & Unwin, 1950.

[14] Of course economists defended free trade in grain also—as perhaps one should say, directly—on the principle of comparative cost, but this more general theory had no special relevance to the politically strategic commodity, food, nor did it necessarily have a prospering beneficiary, that is, a villain.

[15] Mill, John Stuart, *Dissertations and discussions* **5**, Papers on Land Tenure, London, L. J. Parker & Son, 1859. Mill recognized the validity of the landowners' claim to compensation for any differential taxation, and proposed that only future increments of land values be taken by the state.

[16] Lawrence, E. P., *Henry George in the British Isles*, East Lansing, Michigan State Univ. Press, 1957.

[17] Fabian Tract No. 45, 4 edition, 1893. Also *Fabian essays*, 165, 167.

[18] The rate of interest and the laws of distribution, *Quart. Jour. of Economics* **2**: 188–209; also, The rate of interest, *ibid.*, 469–472, 1887–1888.

[19] Stamp, J. C., *British incomes and property*, London, P. S. King & Son, 1916.

probably been declining since 1810. One careful student set the pure agricultural land rents at £6 million at the end of the century.[20]

One could divert attention to urban rents, and the Duke of Bedford made a regular appearance in Fabian publications. Shaw asserted that "town rents have risen oppressively," [21] and Webb was no doubt the author of *The Unearned Increment*, which dilated upon the unearned increment in London, which even his one-sided manipulations raised only to £6 million per year in the twenty-five years up to 1895.[22]

But no amount of literary skill or empirical absent-mindedness can make a Duke of Bedford the arch-villain of capitalism. It was necessary to generalize the indictment, and the attempt was made by both men. They sought to include interest on capital with rent on land, and they denounced the unequal distribution of income. Both criticisms, however, had to be based upon the theory of rent, or a second, independent theoretical basis had to be found for explaining the workings of a capitalistic system. The second alternative was not chosen: it is not only intellectually inelegant, but psychologically ineffective, to use several unrelated theoretical principles to launch a program of reform.

Shaw's endeavors to extend the theory of rent to interest, that is, to identify the returns to capital and labor—were relatively crude. The basic method was that of assertion:

> Colloquially, one property with a farm on it is said to be land yielding rent; whilst another, with a railway on it, is called capital yielding interest. But economically there is no distinction between them when they once become sources of revenue . . . shareholder and landlord live alike on the produce extracted from their property by the labor of the proletariat.[23]

This argument is, in effect, that all capital instruments incorporate land in some degree—the unique location and resources of a railroad or a water company are similar to those of agricultural or urban land. Of course this is true, and "land" is in fact only an abbreviation, in the classical economics, for non-producible, non-human resources. But this element of land hardly dominates the vast mass of capital of an economy, which is both fluid in form and augmentable in quantity, and therefore not obedient to such classical theorems as that the rent of a piece of land increases with economic progress. Shaw's dogmatic assertion that capital produces nothing, and interest is a mere exaction, is simple Marxism, wholly inconsistent with the marginal utility theory of value he professed.

Webb's attempt to identify rent and interest took a more sophisticated form. He generalized the rent theory to cover other distributive shares. If rent is the surplus of the yield of resources on good land over the yield on poor land, can one not also say that interest is the surplus of return of resources (including land) on good capital over the yield on decrepit and obsolete capital?

> "Economic interest" is the amount of produce over and above "economic wages" which is obtained through the use of capital, upon land at the margin of cultivation by the skill of the worst worker employed in the industrial community, or upon the better land with greater skill, after deduction of the economic rent of the land and ability.
>
> Economic interest, as here defined, is expressed by a law similar to the Ricardian law of rent.[24]

Webb argues that most interest is the product of "opportunity and chance." The basic theorem that the rates of interest on various investments tend to equality, is conceded only for "the last increment of capital employed in each case." [25]

It is true that in one sense the rent theory can be generalized to embrace every form of return—Wicksteed demonstrated in 1894 that the rent theory in this sense is analytically equivalent to the marginal productivity theory.[26] But the rent theory, in this sense, is only a device for isolating the contribution of one productive factor to a product, and has no relevance to the crucial assumption of fixity of supplies which gave the Ricardian rent theory its empirical content.

Without this assumption of fixity of supply, there is little substantive similarity between capital and land. Under competition all forms of investment which yield higher than average rates

[20] Thompson, R. J., An inquiry into the rent of agricultural land, *Jour. Royal Statistical Society* **70**: 587 ff., 1907.

[21] Shaw, G. B., The impossibility of anarchism, 9n.

[22] Tract No. 30, 1895.

[23] *Fabian essays*, 19. He subsequently refers to "a form of rent called interest, obtained by special adaptations of land to production by the application of capital . . ." (*ibid.*, 25).

[24] The national dividend and its distribution, in S. and B. Webb, *Problems of modern industry*, 218, 219, London, New York and Bombay, Longmans, Green, 1898.

[25] *Ibid.*, 220.

[26] Wicksteed, Phillip Henry, *Coordination of the laws of distribution*, London, Macmillan, 1894.

of return will attract additional investment until the rate of return is equalized on all, not merely on marginal, increments of investment.

One could seek escape from this conclusion by denying the existence of competition—by arguing that monopolies of various sorts had become so ubiquitous and powerful that most interest was a monopoly return.[27] But Webb placed no emphasis upon such a development. On the contrary, he asserts, that "every development toward a freer Individualism must, indeed, inevitably emphasize the power of the owner of the superior instruments of wealth-production to obtain for himself all the advantages of their superiority." [28]

> But the monopoly of which the democracy is here impatient is not that of any single individual, but that of the class itself. What the workers are objecting to is . . . the creation of a new feudal class of industry, . . . who compete, it is true, among themselves, but who are nevertheless able, as a class, to preserve a very real control over the lives of those who depend upon their own daily labor.[29]

In Webb's scheme, indeed monopoly was not even an evil. In *Industrial Democracy* he presents the famous picture of "higgling in the market." [29] The search of consumers for bargains forces down retailers' prices, retailers in turn are compelled to drive down wholesalers' prices, and so on until the competitive pressure "finally crushes the isolated workman at the bottom of the pyramid." [30] Monopolies are devices to obtain some relief from this unrelenting pressure, and, if they succeed in diminishing competition, they treat their workers a *little* better. Here monopoly is a modestly benevolent phenomenon.

In this later volume, Webb alleges a further resemblance between land and capital, but it appears to be *ad hoc:* the supply of savings is held to be virtually independent of the rate of interest.[31] But even if this were wholly true, it would not yield the Ricardian conclusion that the capitalist necessarily benefits from every advance of society —what happened would depend upon the supply of capital over time.[32]

Both Webb and Shaw seek to extend the rent theory even farther, to include the "rent of ability" of superior workers.[33] It is a most unusual feature of Fabian socialism that it attacked large labor incomes as well as property incomes. Much of the superior earnings of professional men and artists was attributed to the unequal distribution of income, which allowed a few men to prosper by catering to the whims of the rich. Even more was attributed to the "monopoly" of education by the children of property-owning families. The chief remedy proposed was widespread education (which was already developing rapidly), supplemented if necessary by a progressive income tax.

In one respect the analogy between labor and land was closer than that between capital and land: there are natural differences in quality of both men and acres, unlikely to be eliminated under any social system. But the fixity of supply of land finds no parallel in population, and both classical economists and almost all socialists concurred in the absence of any secular increase in wage rates as a society progressed economically. The Fabian doctrine of rents of ability tended both to blur the ethical differences they alleged between wages and other incomes, and to alienate the professional classes from their doctrines, and perhaps these are the reasons this aspect of their theory received little elaboration or emphasis. Still, the socialization of the labor force was latent in this doctrine, and the various hints of compulsion which were dropped [34] are in keeping with the anti-democratic tendencies both Shaw and Webb made explicit in later years.

The final indictment of capitalism, that it generated a cruel and inhuman distribution of income, was no doubt the most influential. The street corner orator was especially clear in Shaw when he reached this theme:

[27] It would still not be a rent, in the sense that a reduction in this monopoly return brought about by price fixing or taxation would usually lead to a reduction in the monopolist's output.

[28] *Problems of modern industry*, 237. Also *Fabian essays*, 55.

[29] Part III, Ch. II.

[30] *Op. cit.*, 671.

[31] *Ibid.*, 622 ff.

[32] In *Capital and land* (Fabian Tract No. 7, 7th ed., 1908) a wholly different criticism of private receipt of interest is advanced by Sidney Olivier. It is to the effect that savings are invested to form concrete capital goods, and when these goods wear out, the goods which replace them are in some mysterious sense not due to the original savings. Shaw uses this argument also in *The Intelligent woman's guide to socialism and capitalism*, Garden City, New York, Garden City Publ. Co., *ca.* 1928; London, Constable, 1928.

[33] Thus Webb says the Ricardian theory must be extended "to all the instruments of production, as well as to the varying efficiencies of every kind of human labour," The Rate of Interest, *loc. cit.*, 472; for Shaw's statement, see *Fabian essays*, 183 ff.

[34] For example, *Fabian essays*, 55–57. See also E. Halévy, *L'ère des tyrannies*, 217–218, 3rd ed., Paris, Gallimard, *ca.* 1938.

A New York lady, for instance, having a nature of exquisite sensibility, orders an elegant rosewood and silver coffin, upholstered in pink satin, for her dead dog. It is made: and meanwhile a live child is prowling barefooted and hunger-stunted in the frozen gutter outside.[35]

This sort of diatribe—adorned by frequent references to such strange capitalistic institutions as compulsory prostitution—must have been especially effective at the lips of one of the most formidable debaters of his time.

The denunciation of inequality may be viewed as an ethical judgment, to be accepted or rejected according to one's taste. This is not a very useful view, however, for equality is not a basic ethical value in any important western philosophy.[36] One surely wishes to distinguish nominal from real inequality: the difference between the average earnings of a twenty-five-year-old lawyer and one twice his age is devoid of ethical significance. If policies are to have purpose and effectiveness, one must isolate the sources of inequality, be they education, natural inheritance, luck, thrift, property inheritance, or a particular government policy. One should seek to quantify the subtle and ambiguous concept of inequality, discover whether it is increasing or decreasing, and invent and analyze alternative methods of dealing with objectionable inequality.

Shaw and Webb discharged a portion, but only a very small portion, of the duties of a responsible proponent of an egalitarian program. Shaw advanced, in fact, three basic arguments for equality of income. The first was simply that there is no other objective basis for distribution:

> Now . . . suppose you think there should be some other standard applied to men, I ask you not to waste time arguing about it in the abstract, but bring it down to a concrete case at once. Let me take a very obvious case. I am an exceedingly clever man. There can be absolutely no question at all in my case that in some ways I am above the average of mankind in talent. You laugh; but I presume you are not laughing at the fact, but only because I do not bore you with the usual modest cough. . . . Now pick out somebody not quite so clever. How much am I to have and how much is he to have? I notice a blank expression on your countenances. You are utterly unable to answer the question. . . .
>
> It is now plain that if you are going to have any inequalities of income, they must be arbitrary inequalities.[37]

This argument cannot be taken as seriously as it was given. Equality is an unambiguous rule of distribution only when it is applied as unmeaning arithmetic, giving equal sums of income to the day-old baby, the adult worker, and the jailed felon. Conversely, a competitive market does determine, not how much more clever Shaw was than contemporary dramatists, but how much more he produced of what people desired.

Shaw's second argument for equality was professedly "economic":

> . . . if you allow the purchasing power of one class to fall below the level of the vital necessities of subsistence, and at the same time allow the purchasing power of another class to rise considerably above it into the region of luxuries, then you find inevitably that those people with that superfluity determine production to the output of luxuries, while at the same time the necessities that are wanted at the other end cannot be sold, and are therefore not produced. I have put it as shortly as possible; but that is the economic argument in favor of equality of income.[38]

That an unequal distribution of income will lead to an unequal distribution of consumption, however, is not an argument for equality. If one believes, as almost everyone always has, that no living person should be denied subsistence (and perhaps much more), then he should favor raising consumption of some people, but the belief does not lead to egalitarianism.

Shaw's final argument for egalitarianism was eugenic; economic barriers prevented marriage from taking place whenever biological urges dictated:

> Just consider what occurs at the present time. I walk down Oxford Street, let me say, as a young man. I see a woman who takes my fancy. I fall in

[35] *Fabian essays*, 21.

[36] For every Shavian parable on the inequities of inequality, one could contrive a counter-parable on the inequities of equality; e.g., "Dr. John Upright, the young physician, devoted every energy of his being to the curing of the illnesses of his patients. No hours were too long, no demand on his skill or sympathies too great, if a man or child could be helped. He received £2,000 net each year, until he died at the age of 41 from over-work. Dr. Henry Leisure, on the contrary, insisted that even patients with broken legs be brought to his office only on Tuesdays, Thursdays, and Fridays, between 12:30 and 3:30 P. M. He preferred to take three patients simultaneously, so he could advise while playing bridge, at which he cheated. He received £2,000 net each year, until he retired at the age of 84."

[37] The case for equality, in *The socialism of Shaw*, 53–54.

[38] *Ibid.*, 60.

love with her. It would seem very sensible, in an intelligent community, that I should take off my hat and say to this lady: "Will you excuse me; but you attract me very strongly, and if you are not already engaged, would you mind taking my name and address and considering whether you would care to marry me?" Now I have no such chance at present. Probably when I meet that woman, she is either a charwoman and I cannot marry her, or else she is a duchess and she will not marry me.[39]

Of this argument I will say only that Shaw was utterly sincere, and that it did not become a Fabian article of faith.

Webb based his analysis of inequality much more directly upon the rent theory.

> Nor is there any doubt or dispute as to the causes of this inequality. The supersession of the Small by the Great Industry has given the main fruits of invention and the new power over Nature to a comparatively small proprietary class, upon whom the mass of the people are dependent for leave to earn their living. When it suits any person having the use of land and capital to employ the worker, this is done only on the condition that two important deductions, rent and interest, can be made from his product, for the benefit of two, in this capacity, absolutely unproductive classes—those exercising the bare ownership of land and capital.[40]

That "the most virtuous artisan cannot dodge the law of rent" [41] is the anchor of this indictment of inequality.

The net effect of rent upon the distribution of income among families in late Victorian England is not known. If it parallels modern American experience—and there should be at least a family resemblance—the wealthiest classes received a much more than proportionate share of rents, and so, too, did the lowest income classes, and the intermediate classes received a less than proportionate share.[42] Elimination of rents, therefore, would reduce income inequality at the top of the income distribution and increase it at the lower end. But no matter what one did with rents, the distribution of income would change little.

Only if one includes interest on capital are the possibilities of redistribution large, and here the rent theory was of no avail: one must contrive an entirely new reason for rejecting private property in capital. Of course such reasons can be contrived, but the Fabians did not meet this demand.

3. CONCLUSION

Reformers seldom proportion criticism carefully to evil, but the Fabians must be among the most extreme in their concentration upon a minor and uncharacteristic aspect of capitalism as its major flaw. There is a non-functional income to productive resources in fixed supply, and it is roughly —but only very roughly—approximated by rent of land, but it was already trifling in their time and it was on balance (including urban rents) declining relative to national income.

One expects blind slogans and high emotions to carry a mass movement, but the particular clientele of the Fabians was the educated class of Great Britain. That they were as successful as they are reputed to be,[43] is suggestive of a proposition for which I believe there is much support: that social problems are the creation of the "intellectual." The intrinsic importance of a complaint against a social system, as judged by later opinion, has little to do with its effectiveness in shifting opinion. If enough able and determined men—and the number in the Fabian group was almost unbelievably small—denounce and denounce again a deficiency, that deficiency becomes grave.[44]

It is less unusual of the Fabian theoreticians that they were not good economists—popular reformers have seldom been. In economics Shaw was merely a clever dilettante. No short sentence could do justice to Webb's large talents, but they did not include a strong command of economic theory. Their limitations prevented them from constructing a coherent program of economic reform, but it apparently did not decrease the effectiveness of their criticisms of the existing order. Had the leading economists of the period subjected this literature to critical review—instead they simply ignored it—I doubt whether the control over land use would have been an element of modern English socialism.

[39] *Ibid.*, 63–64.

[40] The difficulties of individualism, in *Problems of modern industry*, 235–236.

[41] *Ibid.*, 260.

[42] See D. Gale Johnson, Rent control and the distribution of income, *Proceedings*, American Economic Association, 568–582, May, 1951.

[43] Their influence is a difficult question, to which no established answer has yet been given. They are credited with much long run influence, but little immediate impact, by H. Pelling, *The origins of the Labour Party, 1880–1900*, London, Macmillan, 1954; New York, St. Martin's Press, 1954.

[44] Of course, to the possibly decisive extent that the distribution of income, rent theory aside, was the successful element of their indictment, they were only one of many streams of nineteenth century criticism, and their influence must be assessed correspondingly lower.

DR. JOHN EVANS, U. S. GEOLOGIST TO THE OREGON AND WASHINGTON TERRITORIES

ERWIN F. LANGE

Professor of General Science, Portland State College

AMONG the obscure and unrecognized scientific explorers of the northwestern part of the United States is Dr. John Evans, United States geologist, who devoted his life to science and to the service of his country. In spite of his valuable scientific discoveries and outstanding contributions to his nation, no complete account of his life and work has yet been written and his important monograph remains lost and unpublished.

John Evans was born February 14, 1812, in Portsmouth, New Hampshire, the son of Richard Evans, an associate justice of the New Hampshire supreme court. John Evans received his early education in the schools of Andover, Massachusetts, and devoted several years to mercantile pursuits in that state. At the age of nineteen he moved to Washington, D. C., then commonly known as Washington City, where he was employed as a clerk in the general post office department for eight years. While residing in Washington, he married Sarah Zane Mills on May 16, 1835. She was the daughter of Robert Mills, engineer, and architect of the Washington monument. Three sons and one daughter were born into the Evans family. In 1839 Evans and his family moved to St. Louis where he studied medicine and received his M.D. degree from the St. Louis University Medical Department. At the commencement held March 3, 1854, the Medical Department of the same school conferred on him the "Honorary Degree." [1]

Evans belonged to three scientific societies: the Boston Society of Natural History, the Academy of Natural Sciences of Philadelphia, and the Academy of Science of St. Louis.

In 1847, Dr. David Dale Owen who was appointed United States geologist to conduct a geological survey of Wisconsin, Iowa, Minnesota, and part of Nebraska, chose as sub-agents Dr. John Evans and Dr. B. F. Shumard. The work of Evans soon attracted the attention of Owen so that great confidence was placed in his ability. In 1849 the commissioner of the general land office directed the survey to be extended further into Iowa than had originally been intended. Sub-agent Evans was selected to proceed up the Missouri River in advance of the main corps, and deposit provisions at suitable points. He was further instructed to carry on the examination of the Cretaceous and Tertiary formations west of the Missouri as time would permit. He was aided in his travels by being able to use the facilities of the American Fur company in both land and water transportation and was thus able to extend his observations over a larger area than had been anticipated.

The climax of his exploration occurred when he reached the most curious unexplored region, the country of the Bad Lands (frequently called the Mauvaises Terres by the French Canadian trappers), lying high on the White River. Apparently the area was explored in the heat of the summer sun, for the excitement of the discovery was described by Owen: [2]

> The drooping spirits of the scorched geologist are not permitted, however, to flag. The fossil treasures of the way, well repay the sultriness and fatigue. At every step, objects of highest interest present themselves. Embedded in the debris, lie strewn in great profusion, organic relics of extinct animals.

Of Evan's work here in the Bad Lands, Dr. Shumard wrote: [3]

> During this tour Dr. Evans made an excursion to the celebrated Mauvaises Terres of Nebraska, and he has credit of having been the first scientific explorer who visited those vast cemetaries of extinct animals which have excited so much interest among scientific men. And such was his energy that almost un-

[1] The date on which John Evans received his M.D. degree cannot be ascertained. He was frequently referred to as Dr. John Evans during the period 1851–1853. *The Annual Catalogue and Announcement of the Medical Department of the St. Louis University* for the 1854–1855 session states: "The Honorary Degree was conferred on Dr. John Evans of Washington City."

[2] Owen, David Dale, *Report of a Geological Survey of Wisconsin, Iowa, and Minnesota*, 197, Philadelphia, 1852.

[3] Shumard, B. F., Dr. John Evans, *Trans. Acad. Sci. St. Louis* **2**: 162, 1861–1868.

aided, he gathered in a few weeks the extensive collection of fossil bones which furnished Dr. (Joseph) Leidy the material for his splendid memoir on the "Extinct Fossil Fauna of Nebraska."

One account of the day reported that the fossils from the Bad Lands created such interest among European geologists that they immediately sent an eminent member of their group to this country to study the collection and to examine the locality from which it was taken.[4]

During the same excursion Evans also collected a "large and elegant" series of Cretaceous fossils which were described by Owen and Shumard in the report of the survey. In this paper, describing new species of Crinoidea, one was named *Megistrocrinus evansii* in honor of the finder. Of this the authors wrote:[5] ". . . for the magnificent specimen figured we are indebted to John Evans of the United States Geological Corps, and we have dedicated it to him."

Dr. Shumard further wrote concerning the work of Evans with the survey:[6]

> The notes made by Dr. Evans on the geological features of that remarkable region (Bad Lands) were likewise of the highest interest, and from these Dr. Owen has compiled one of the most instructive and beautiful chapters to be found in his valuable report. In addition to his scientific duties, Dr. Evans had almost exclusive control of the business department of Dr. Owen's survey, which of itself involved an immense deal of labor. The satisfactory manner in which he discharged these onerous duties, often in the midst of disheartening privations and even danger, commanded the highest esteem and confidence of Dr. Owen and his associates, while his goodness of heart, uniform courtesy, and self-sacrificing disposition, secured to him their warmest friendship.

In March, 1851, on the recommendation of Dr. David Dale Owen, Dr. John Evans was appointed "to institute Geological researches on the main lines of the public land surveys about to be commenced in Oregon" by the Secretary of the Department of Interior. Evans had already proposed to the Department the extent of the survey, and the instructions which he received were largely those of his own suggestion. According to these instructions, Evans was to carry on a geological and mineralogical survey west of the Cascade Mountains beginning on the main lines of the public land surveys. He was to determine barometrically the elevation of mountain ranges, peaks, and spurs. He was instructed to keep a detailed register of barometric observations and of geological and mineralogical researches together with the inferences drawn from the work. He was to determine the latitude and longitude of various points, particularly where the Base line was to intersect the principal meridian, and the latitude and longitude of the mound to be erected on the summit of the Cascade Mountains which was to be the legal terminus of the Base line.[7]

Before beginning his overland journey, Evans wrote the General Land Office on March 8, 1851:[8]

> In the overland trip, I shall require the services of one half breed voyageur and two pack mules; and probably shall find it indispensable to retain them in Oregon for transportation of instruments, supplies, etc.
>
> Mr. Preston (Surveyor General of the Oregon Territory) I believe has been furnished with two chronometers. I have already a good sextant, artificial horizon, etc., and a good barometer tube, 1 spirit level, two thermometers in tin cases, a compass, and a good telescope. The latter I have found more useful on the prairies as a defense from hostile Indians, than rifles and revolvers, for it enables you to avoid them.

The salary for Evans for the survey was listed at $6.00 per day to be taken from the $3,500.00 appropriated for the project. He stated he would undertake the survey for the funds available and promised to practice the most rigid economy. However, if the expenses proved to be greater than anticipated he hoped that the government would make up the deficiency. Dr. Evans was also hopeful that he might find a wagon pass through the northern Rocky Mountains shorter than the famous Oregon Trail. On the survey

[4] See the account of Dr. John Evans in *The American Annual Cyclopaedia of the year 1861* 1: 287, New York, D. Appleton & Co., 1872.

[5] Owen, D. D., *op. cit.* Appendix of the Report, 587–588.

[6] Shumard, B. F., *op. cit.* 2: 163.

[7] The writer recently acquired a map of a portion of the Oregon Territory completed by the Surveyor General's Office at Oregon City, Oregon, on October 21, 1852. This map shows the extent of the land surveys and lists the Base line as 45° 31′ 15″ north latitude and the meridian line now known as the Willamette meridian as 122° 30′ 26″ west longitude as "found by Dr. John Evans Sept. 15th, 1852." The accepted values at present for these readings are respectively 45° 31′ 10.83″ and 122° 44′ 33.55″.

[8] Dr. John Evans' letters and reports from which these excerpts are taken, concerning the Geological Survey of the Oregon Territory, are today on file at the National Archives, Washington, D. C.

Evans was again to be aided by Dr. F. B. Shumard who decided to go to Oregon by way of the Isthmus of Panama.

The difficulties and delays of the early weeks of travel have been described in a report of Evans written in July, 1851, from Milk River some two hundred miles above the Yellowstone River: [9]

> In addition to delays and other annoyances on the trip, I might perhaps mention the presence of cholera on board for three weeks after leaving St. Louis. Nearly one half of the passengers were attacked, and fourteen out of one hundred died—every seventh man. Those who remained free from the disease were completely prostrated by the foul atmosphere, and constant attendance on the sick, day and night. A Catholic missionary among the Indians, . . . fell a victim to his benevolent efforts to aid the sufferers. Completely worn out he sunk under the disease in a few hours. He was attacked and sent for me at 3 o'clock in the morning, and at 9 o'clock the same morning he was draped in his priestly robe for the grave—peace to the ashes of a good man.
>
> The principal man of those engaged to accompany me over the mountains was killed a short time since by accident. I have since engaged two others from the Fur Company and three pack mules.

Evans wrote his next report on October 21, 1851, from Oregon City, Oregon. In this detailed letter he described his route across the continent and wrote of his many difficulties and hardships. Of his route he wrote as follows: [10]

> After my hasty communication from Yellow Stone, I proceeded up the Missouri River travelling alternately on the north and south side, and making occasional excursions to the interior.
>
> We entered the "Black Foot Pass" by Medicine or Sun river to the heart of the mountains, and then proceeded in a Northwest direction to the British Line, a distance of three degrees of latitude, examining every gap in the mountains that had the appearance of a pass. Returning south I then crossed over an elevated range of mountains to Flat Head Lake; then took a south easterly direction along the beautiful, and in many places, fertile valley bordering the Western base of the Rocky Mountains.
>
> After examining the valley of St. Mary's or Bitter Root River I went up the most Western, or Lou Lou fork of the Bitter Root, crossed the St. Mary's, Spokane, and Kooskooskie range of mountains, all meeting at this point, and forming a congregation of mountain peaks, extending 130 miles. Passing into the valley of the Nez Perces, we followed the general direction of the Kooskooskie or Clear Water River to its junction with Snake River crossing the Walla Walla, John Day, and Fall River to the Cascades.
>
> Along the whole route I have been enabled to collect and preserve specimens sufficient for analysis; and to make a general geological section from the Cretaceous formations of the Missouri, to the Pacific Ocean.

The wanderings of Evans led him through many uncharted forests and mountains never before traversed by civilized man. His difficulties and hardships were vividly described in the same letter:

> I travelled sometimes alone, sometimes with two half breeds or Indians, employed from village to village. I dispensed with a tent and many of the usual comforts of voyaging, taking only sufficient provisions of the plainest kind to sustain life; in order to economize the means of transportation for my specimens; having literally thrown away bread to preserve stones! On one occasion I passed two days without food or water, toiling from early dawn till night in climbing a steep mountain side. Another time I was obliged to kill my favorite dog which had followed me in all my wanderings. Two of my horses were left to die in the main chain of the Rocky Mountains, and I was obliged to leave two more in the mountains on this side; and my poor mules were so galled and jaded that it was pitiable to behold. Those who have crossed these mountains with experienced guides, following plain Indian trails, with tents and other camping equipage, with escort, servants, and abundant supplies, may perhaps form some idea of what the same journey would be, without companion, guide, or camp equipage, not through passes but over rugged mountains they see rising on either hand, to say nothing of the more inaccessible ranges further north. . . . On one day I crossed a mountain torrent 18 times, over large and slimy boulders; and jumped my mule over *897* fallen trees, amidst tangled thickets, on the steep mountain side. These logs were too large for my animals to step over.

Evans had already found that it was impossible to complete the survey for the $3,500 appropriated by the government. In his letter from Yellowstone he informed the Department that he planned to stay in Oregon the following year and that the cost of the survey would require $10,000 which he hoped would be appropriated. In his next letter he stated that the cost of the survey to January, 1852, would cost $5,000 because of the uncertain living conditions in a wild and uncivilized land. Of his financial difficulties he wrote the Department: [11]

[9] *Ibid.*
[10] *Ibid.*
[11] *Ibid.*

Very few of the government salaries in the Territory, if any, will pay expenses. The lowest board that can be obtained is $12 per week, and transient board $4 per day. Washing $4 per dozen—horse hire $3 per day—horse keeping $2 per day—hire of pack men from $2.50 to $3 per day—rent of room for arrangement and preservation of specimens, and making analysis of ores, etc., $20 per month; and other expenses in proportion.

On November 18, 1851, the *Oregon Spectator*, the first newspaper published west of the Rocky Mountains, announced that Dr. Evans had arrived in Oregon City several weeks prior. Here he met Dr. Shumard who had arrived in August but had spent the intervening time exploring the Willamette Valley, and the Umpqua Valley in southern Oregon. The account also referred to the interesting collection of fossils which Dr. Shumard had already gathered in Oregon.

FIG. 1. The Willamette stone marking the intersection of the base line and the Willamette meridian, located just off Northwest Skyline Drive, Portland, Oregon.

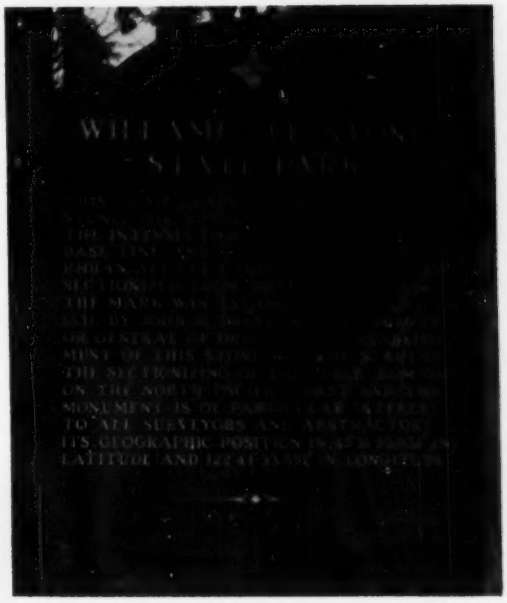


FIG. 2. The Willamette Stone State Park marker on Northwest Skyline Drive, Portland, Oregon.

The details of the geological survey in Oregon are to be found in the letters of Evans written at regular intervals to the General Land Office in Washington. During the winter of 1851 and the spring and summer of 1852 Evans and Shumard carried on the work of the survey beginning along the Willamette meridian and the Base line as they had been directed. Their work was primarily confined to the territory south of the Columbia River. The letters relate a story of hardship and difficulty in respect to every phase of the work. In June, 1852, Evans reported that he had financed the survey for eight months, and in August of 1852 he made an appeal for an immediate advance of $1,500 to continue the work and reported that he had already borrowed $4,218 to carry on to that date. The work was further hindered by a lack of chemical apparatus and reagents needed for analytical work, and the absence of scientific reference books was keenly felt. However, a large number of samples of soils, coal, minerals, and fossils were forwarded and the work was continued, and received the approval and commendation of the populace of

the territory. In August of 1852 Dr. Shumard returned to Washington while Dr. Evans remained to explore the territory north of the Columbia River. At the end of 1852 Dr. Evans returned to Washington by way of the Isthmus of Panama.

A paper describing new fossils collected by Evans from the Oregon and Washington Territories was published by Shumard in the *Transactions* of the Academy of Science of St. Louis.[12] A survey of this article indicates that fossil species were collected at such divergent points as Port Orford and Coos Bay on the southern Oregon coast, between Salem and Oregon City in the Willamette Valley, near Astoria at the mouth of the Columbia River, and as far north as Vancouver Island. It is quite likely that some of these were collected by Evans on his next trip to the Oregon Country. This paper was considered of sufficient scientific importance to be republished in 1909 in the *United States Geological Professional Paper* 59.

Dr. Evans next served science and his country as geologist of the railroad survey along the northern route. In March of 1853 Congress formed the Washington Territory from the northern half of the Oregon Territory. One of the first acts of the new President of the United States, Franklin Pierce, was to send to the Senate the nomination of Major Isaac I. Stevens as governor of the Washington Territory. During that same month Governor Stevens was placed in charge of the exploration of the northern route for a railroad by Secretary of War, Jefferson Davis. In making preparations for the survey, Stevens wrote the Secretary of War on March 25 that the distinguished geologist, Dr. John Evans, had been retained as geologist of the survey. In his report of the survey Governor Stevens further referred to the appointment of Evans:[13]

> Dr. John Evans was appointed geologist of the exploration. Since 1851 he had been in the services of the Department of Interior, making a geological survey of the Oregon and Washington Territories, and it was believed that the public would be best promoted by securing his cooperation with the geological portion of the exploration. He has passed over portions of the route, and, whilst preparing for the expedition, he furnished me with information in regard to the general course of Milk river, of passages from the Missouri to Bitter Root valley, of the valley itself, and of the route thence of Lewis and Clark, by the Lou-Lou, or Travelers Rest creek, to Fort Walla Walla. He drew instructions to guide in making the collection of the exploration, and was himself provided with funds deemed, in his judgment, necessary to make the trip across the continent in connection with the work of all parties of the exploration. It seemed that thus the most ample and effective means were taken, both to make the geological results of the exploration complete, and to advance the general services on which Dr. Evans was employed by the Interior Department.

Not only were the services of Evans valuable to the survey because of his knowledge of the territory to be explored but also for the directions for observations to be followed by the geological workers of the expedition. Since the survey was to be about two thousand miles long and two hundred and fifty miles wide, extending from St. Paul in Minnesota to Puget Sound in the Washington Territory, the cooperation of a large number of scientific and non-scientific workers was required. Clear and detailed instructions provided by Evans for the conduct of the geological survey were a portion of the report to the Secretary of War and were published in House Executive Document 46 of the 1st session of the 33rd Congress (1854).

Few details of the work of Evans on the survey were published.[14] However, Governor Stevens

[12] Shumard, B. F., Description of new fossils from the Tertiary formation of Oregon and Washington Territories and the Cretaceous of Vancouver Island, collected by Dr. John Evans, *Trans. Acad. Sci. St. Louis* **1**: 120–125, 1858. See also: Lesquereux, Leo, On some fossil plants of recent formations, *Amer. Jour. of Science and Arts* (second series) **27**: 359–363, 1859. Lesquereux, Leo, On fossil plants collected by Dr. John Evans at Vancouver Island and at Bellingham Bay, Washington Territory, *Amer. Jour. of Science and Arts* (second series) **28**: 85–89, 1859.

[13] Stevens, I. I., *Report of the exploration and survey to ascertain the most practicable and economic route for a railroad from the Mississippi River to the Pacific Ocean* **12**: 33, book 1, Washington, 1860.

[14] Merrill, G. P., History of American geology, *Smithsonian Report for 1904*, Washington, 1906. On page 455 the author mentions that Dr. Evans prepared two reports on the geology of the northern route and turned them over to I. I. Stevens. Both reports were lost in transit from the Pacific Coast to Washington. This is verified by Stevens, *op. cit.* **1**: 230 and **1**: 282. The first report included the geology of the route south of the Missouri and Yellowstone Rivers and between the Milk and Missouri Rivers while the second report dealt with the geology from Fort Benton to the Lower Columbia. Owing to this loss, no geological report of the northern route is included in the voluminous reports of the railroad surveys.

Merrill's further statement that Evans was working on such a report at the time of his death and that such a

in his report of the expedition made several comments of his activities. Evans was directed to examine the country on the southward part of the survey and was given special directions to explore again the "Mauvaises Terres" and was to meet Governor Stevens at Fort Union and Fort Benton.

On September 5, 1853, Governor Stevens reported:[15]

> Dr. Evans also arrived today in fifteen days from Fort Union. He took the route between the Milk and Missouri rivers, passing through the Bear's Paw, and reports having found many delightful, well grassed, and arable valleys in this mountain, and that indeed generally, between these two rivers from their junction, the country was exceedingly well adapted to voyaging; water, grass, and wood, to be found at short distances, and much of the land he represents as being arable. The Doctor made a large collection in the Mauvaises Terres, and made quite a circuit on his way to Fort Union, finding a large extent of country between Fort Pierre and Fort Union, south of the Missouri, a prairie country instead of a country of hills and mountains as had been generally supposed; grass not only good, but the country emphatically cultivable.

Fossil collections from this trip resulted in two papers on new fossils by Evans and Shumard published in the *Proceedings* of the Academy of Natural Sciences of Philadelphia.[16, 17]

On October 22, 1853, the *Portland Oregonian* announced that Dr. Evans had arrived in Portland. The account mentioned that Evans had left Governor Stevens in the Blackfoot country, a hundred miles east of the Rocky Mountains. At this point the expedition was divided into two parties in order to examine two different routes across the mountains. In summarizing the activities of the railroad survey, a somewhat detailed itinerary of Evans was included:

> Dr. Evans has traveled a different route from that taken by Governor Stevens, for the purpose of making as extended a reconnaissance as practicable. His route has been south of the Missouri. Leaving that river at Fort Pierre, he proceeded south to the head waters of the White river, thence along the base of the Blackhills, on the high table land bordering these mountains to Yellow Stone—crossing the Missouri, he ascended the same to Milk river; and from thence midway between the Missouri and Milk rivers to the Blackfoot Pass, through which he came on to the Bitter-root river, following up the stream to the Nez Perce trail; crossing the Bitter-root Mountains, and through the Nez Perce Country to Walla Walla and the Dalles of the Columbia. Dr. Evans is fully satisfied that this route is perfectly feasible for a railroad.

In completing the report of the expedition Stevens praised the work of Evans:[18]

> I will close this communication by advertising, in terms of highest commendation, to the ability with which Dr. John Evans has managed the geological portion of the work, and the great contribution he has personally made in collections, and in developing the geography of the country travelled over by him.

At the conclusion of the railroad survey, Evans remained in the Pacific Northwest to continue the geological survey of the Oregon and Washington Territories which was begun in 1851. Little has been known of his activities during the next several years. However, he kept a day by day diary of his activities for the years of 1854, 1855, and 1856. While the extant portion of the diary is very fragmentary, it does indicate that a large part of the Oregon and Washington Territories was again explored. From Fort Vancouver on the Columbia he started for Puget Sound but the details of this journey have been lost. Next he traveled up the Columbia to Walla Walla and then to Spokane and only part of the return journey is recorded. His final trip in 1856 was to Coos Bay and Port Orford on the southern Oregon coast. He returned to Oregon City by way of the Umpqua and the Willamette Valleys. On August 30, 1856 the *Oregon Argus* (Oregon City) recounted:

> Dr. Evans, the efficient U. S. geologist who has been engaged for several years in making a geological survey of Oregon and Washington Territories, will

report was forwarded to the Government Printing Co. in 1861 seems in error as this latter report was undoubtedly his geological report of the Oregon and Washington Territories which has also become a lost document. Since the volume of the railroad surveys dealing with the northern route was published in 1860, it seems unlikely that Evans would be working on it in 1861. The diary of Dr. Evans for the period July to September, 1853, is on file at the Library of Congress.

[15] Stevens, I. I., *op. cit.* **12**: 104.

[16] Evans, John, and B. F. Shumard, Description of new fossil species from the Cretaceous formation of Sage Creek, Nebraska, collected by the North Pacific Railroad expedition, *Proc. Acad. Nat. Sci. Phila.* **7**: 163–164, 1854.

[17] Evans, John, and B. F. Shumard, Description of new fossil species from the fresh water Tertiary formation of Nebraska, collected by the Northern Pacific Railroad expedition, *Proc. Acad. Nat. Sci. Phila.* **7**: 164–165, 1854.

[18] Stevens, I. I., *op. cit.* **1**: 159.

leave for home on the next steamer. His museum of curiosities gathered on this coast will make a valuable acquisition to the collection already at the capital (Smithsonian).

The next few years of Evans' life seem to have been spent in Washington, D. C., preparing a report for Congress on the geology of the Oregon and Washington Territories. As early as December 6, 1852, a communication from Dr. Evans was read in the House of Representatives along with a resolution for continuance of the work. Evans reported at that time that a reconnaissance had been made from the Rocky Mountains to the Pacific Ocean, including 150 miles north and south along the western base of the Rocky Mountains, and from the northern boundary of California to Puget Sound. The report listed a number of minerals which had been found in this territory and the probable economic value of these minerals. Included in the list were limestone, coal, lead, iron, copper, clay, and gold. In concluding the report he wrote:[19]

> Eastern Oregon has been generally much underrated . . . the districts in that region, afford fine pasturage, and many beautiful sites for agricultural communities. Sufficient information has already been obtained, to render it probable, that many portions of Oregon, will prove rich in useful minerals, and its soil for many agricultural products is unsurpassed. It also affords a rich and interesting field for scientific research.

This report of Evans was included by John P. Gaines, Territorial Governor of Oregon, in his message to the Territorial House of Representatives on December 11, 1852.

However, Evans wanted to make a more complete report and was desirous of again visiting the Northwest. His appointment as geologist to the railroad survey gave him this opportunity. After returning to Washington late in 1856 he worked diligently to finish the report and hoped for its publication by Congress. For reasons not entirely clear, Congress hesitated both to pay Evans for his work and to publish his document. The correspondence of General Joseph B. Lane, Oregon Territorial delegate to Congress, contains a number of letters from prominent people familiar with the survey pointing out the importance of Evans' work and soliciting Lane's help in having Congress make the necessary appropriation. Finally in March, 1855, an appropriation of $23,560 was voted by Congress for "expenses incurred by Dr. John Evans in geological explorations in Oregon, Washington, etc., and for completion of these explorations."[20] No action was taken on publication of the report.

On December 7, 1857, George L. Curry, Territorial Governor of Oregon included the following in his message to the legislative assembly of the Oregon Territory:[21]

> You are doubtless aware that John Evans, a gentleman of high scientific attainments, has been engaged for several years, under direction of the General Government, in making a geological reconnaissance of this and the Washington Territories. With the character of his exploration, and the zeal and energy with which they have been conducted, you are likewise familiar. The publication of his report, which is now completed, ought to be a matter of considerable interest, especially to the two territories, for we may confidently calculate that he has collected much valuable information which will be of importance to the growing interests of this section of the Union. It is desirable that the fruits of his investigations should be made known to the world. I have no doubt that the passage of a joint resolution, by your respective bodies, instructing our delegate to request Congress to make the necessary appropriations for the early publication of the work, might be of service in accomplishing that end. The commissioner of the General Land Office estimates the requisite amount for this purpose at $21,400. He considers the labors of Dr. Evans "as eminently useful in a scientific point of view, as well as subserving the interests of the Pacific Coast, by indicating the localities of the country possessing mineral and agricultural wealth."

Such a resolution was immediately passed by the two houses of the Oregon Legislature and forwarded to Oregon's delegate to Congress.

In December of 1857 a memorial was adopted by the Academy of Science of St. Louis urging Congress to publish the report on the Geological Survey of Oregon and Washington Territories.

On May 13, 1858, United States President James Buchanan officially forwarded the report of Evans to Congress through James L. Orr, Speaker of the House of Representatives. His brief message read:[22]

> I herewith transmit, to be laid before the House of Representatives, the letter of the Secretary of In-

[19] Evans, John, Geological survey of Oregon, *The Columbian* Olympia (Washington Territory), January 22, 1853.

[20] U. S. Congress, 33rd, 2nd Session, *Congressional Globe* (appendix) **24**[2]: 387, Washington, March, 1855.

[21] *Oregon Statesman*, December 15, 1857.

[22] Richardson, J. D., *Messages and papers of the presidents* **4**: 3016, New York, 1897.

terior, dated the 12th instant, covering the report, maps, etc. of the geological survey of Oregon and Washington Territories, which has been made by John Evans, esq., United States geologist, under appropriations made by Congress for that purpose.

The report was not published by Congress in spite of strong resolutions to that effect. In 1861 Dr. Charles T. Jackson of Boston to whom Evans had sent many of his samples for chemical analysis wrote that the report was then in the General Land Office in Washington awaiting publication which he hoped would not long be delayed. The final reference to the report is found in a brief sketch of Dr. Evans written by his co-worker, Dr. Shumard, in 1863:[23]

In the prosecution of the geological survey of Oregon and Washington, Dr. Evans was actively engaged for nearly five years, during which period his travels extended over a large portion of those territories. The result of his important labors, embracing a large amount of information relating to the geology, topography, geography and natural history of that interesting portion of the American Continent, are embodied in a large report submitted to, and ordered to be printed by Congress at its last session. It was placed in the hands of the Public Printer only a few days previous to the death of its lamented author, and it is a matter of deep regret, that owing to some formality, its publication must be delayed to await the further action of Congress. It is earnestly hoped that Congress will, at an early period, make the necessary provision for the printing of this important document, the preparation of which has cost such a large expenditure of labor and money, and which promises to be of the highest value to science and the people for whose benefit the survey was ordered.

The report, which many consider to be the outstanding scientific accomplishment of Evans was never published by Congress and is today a lost document.

The last service of Evans to science and his country was that of geologist to the Chiriqui Exploration Expedition to the Isthmus of Panama in 1860. He was appointed to this position by President Buchanan. Again he performed his duty with zeal and diligence. His delightful geological report of the Chiriqui Expedition was published in House Executive Document 41 of the 36th Congress, 2nd session in January, 1861. Dr. Evans was delighted with the healthful Chiriqui climate and planned eventually to move there.

[23] Shumard, B. F., *op. cit.* **2**: 164.

FIG. 3. Port Orford meteorite. Sample in U. S. National Museum weighs 24.3 grams.

However, a few weeks after returning to Washington he died suddenly on April 13, 1861, at the age of forty-nine.

Although the works of Evans attracted the attention of many of his contemporaries in both governmental and scientific circles, one of his discoveries was to create more interest over the years than any of his activities yet mentioned. This was the discovery of the large Port Orford meteorite in the Oregon Territory. Many of the mineral specimens found by Evans were sent to Dr. Charles T. Jackson of Boston for chemical analysis. In opening one shipment of specimens from Oregon, Dr. Jackson was startled to find a piece of iron which he soon learned came from a large meteoritic mass. A fragment of the meteorite was sent to Dr. W. K. Haidinger, international authority on meteorites in Vienna, who verified the meteoritic nature of the specimen and reported it in the proceedings of the Vienna Academy of Science.[24]

Jackson began correspondence with Evans concerning the location of the meteorite and the possibilities of securing it as a museum specimen. Several scientific societies, including the Boston Natural History Society and the Academy of Science of St. Louis, sent memorials to Congress

[24] Haidinger, W. K., Einige Neure Nachrichten Uber Meteoriten, *Sitzber. Wien. Akad.* **41**: 572, 1860, and Meteoreisen von Rogue River Mountain in Oregon, *Sitzber. Wien. Akad.* **44** (II): 29–30, 1861.

requesting appropriations whereby the meteorite might be removed from Oregon and deposited in the Smithsonian Institution. However, Congress becoming embroiled in the Civil War was not much concerned with a distant meteorite. Next, a group of scientific societies proposed to have Dr. Evans lead an expedition to Oregon for the purpose of recovering the large meteorite. Before such an expedition could be organized, Dr. Evans died, so the plan was abandoned. In the hope that some traveler might be interested in looking for the meteorite Dr. Jackson in 1861 published a number of clues to guide searchers.[25] These clues were excerpts taken from letters of Dr. Evans written in 1859 and 1860 from Washington, D. C. In these letters Dr. Evans described finding a large iron mass projecting out some three to five feet from the top of the western face of "Bald Mountain" within thirty or forty miles of Port Orford. Consequently the meteorite has become known as the Port Orford Meteorite. The mountain was further described as being a prominent landmark, seen for a long distance on the ocean, and higher than any of the surrounding mountains.

The clues of Jackson and the report of Haidinger have attracted the curiosity of many people. Both scientists and amateurs have searched without success for the lost meteorite which is considered to be the tenth largest ever discovered. Its mass is estimated at ten or eleven tons. The territory within a thirty or forty mile radius of Port Orford is some of the roughest terrain in Oregon and the hills or mountains in that area are numerous. "Bald Mountain" was a general term applied to various mountains having high grassy prairies. At present there are specific mountains named Bald Mountain and Bald Knob. Evans also pointed out that the mountain was subject to washing from heavy rains and melting snows which might cover up a large portion or all of the exposed meteorite. In the period of a hundred years since Evans visited this area the bald mountains have become grown over with scrubby vegetation so that the clues have become more difficult to follow. A study of Dr. Evans' carefully written diary and a comparison of his geological notes with the Port Orford quadrangle identify his "Bald Mountain" as the present Iron Mountain, which has an elevation of 4,004 feet and is higher than any of the surrounding mountains. Perhaps, as the search continues and all clues are carefully studied this world famous meteorite may yet be recovered.

The passes used by the Great Northern and the Northern Pacific Railroads through the Rocky Mountains in Montana are today useful monuments to this unselfish and determined geological explorer. The Bad Lands of South Dakota, first explored by Evans, are important tourist attractions, and the memory of Dr. Evans on "Bald Mountain" lingers on as the relentless search continues for what could be the largest meteorite yet discovered in the United States.

At his death, the Boston Society of Natural History paused to adopt the following resolution:[26]

> Resolved, that the Boston Society of Natural History have learned with deep regret the decease of their corresponding member, Dr. John Evans, United States Geologist of Oregon and Washington Territories and of the Chiriqui Expedition, and deplore the loss which the country has sustained in the death of this active and enterprising geologist.

[25] Jackson, Charles T., A sketch of the life and scientific services of John Evans, M.D., U. S. Geologist for Oregon and Washington Territories, *Amer. Jour. Sci.* **32**: 313–314, 1861.

[26] *Proc. Boston Soc. Nat. Hist.* **8**: 177, 1861.

REFERENCES

EVANS, RICHARD X. 1935. Dr. John Evans, U. S. Geologist, 1851–1861. *Washington Hist. Quart.* **26**: 83–89.

STEVENS, HAZARD. 1900. The life of Isaac Ingalls Stevens. 2 v. Houghton Mifflin Co. (Published by the Riverside Press, Cambridge.)

AN AXIOMATIZATION OF GENERAL RELATIVITY

RICHARD A. MOULD*

Assistant Professor, New York State University College on Long Island, Oyster Bay, N. Y.

I. THE PROBLEM AND METHOD

It is frequently maintained that the General Theory of Relativity gives scientific support to the thesis that space and time are a manifestation of the properties of material objects. This thesis suggests that physical space-time requires the prior notion of matter, or material object, for its definition and meaning. The philosopher Ernst Mach is widely associated with this view. Mach's influence on Einstein, and the latter's success in giving physics a theory which displays the interdependence of spacial and material concepts is believed to provide a strong endorsement of the view that material objects somehow generate or define physical space. So far as the writer is aware, however, there has been no convincing demonstration that matter does play so important a role in the Einstein theory. Briefly considered, general relativity is a field theory.[1] It deals directly with a four dimensional continuum of points and geometrical notions such as metric and affinity. The major geometrical and kinematical ideas of the theory are normally developed prior to, and without the assistance of dynamical variables such as mass, momentum, and pressure.

Central to those dynamical quantities which do occur in the theory is the matter tensor T_{kl}; and superficially at least, this tensor appears to be defined in terms of geometric-kinematic concepts. It is furthermore meaningful in the formalism of general relativity to speak of a space in which this matter tensor is everywhere zero. The theory, therefore, provides a meaning for the specification of a space devoid of material particles. On the other hand it fails to provide a unique determination of geometry and universal topology in many cases when a matter content is completely specified. If Einstein's general theory does encourage the view that space-time is identically a relation holding between material particles, then these formal features of the theory seem paradoxical.[2,3]

One holding the above view, which we will call the Machian thesis, cannot plausibly avoid questions of formal relevance by arguing that "material object" is intended only to mean "empirical object." An empiricist may certainly insist that physical space is identically the collection of relations which hold between empirically given objects as relata, but then the special features of a logico-mathematical formalism which embodies those relations (viz: the relativity theory) need not concern him. If Einstein's theory is to provide some manner of support for a philosophical thesis of the Machian kind, then the logical structure of that theory must provide decisive elements.

An alternative to Mach's thesis is the warp-in-space idea of Clifford[4,5] which the writer believes it is more appropriate to the sense of general relativity than the view of Mach. According to Clifford, the dynamical properties of matter are to be understood in terms of local distortions or curvatures in space. Material particles are given their meaning and definition by reference to geometrical notions which are assumed to have a prior status in physical science.[6] On this thesis, it is reasonable to imagine a space devoid of material particles, and to expect the major

* The author wishes to express his indebtedness to Professor Henry Margenau, Yale University, for stimulating and guiding the performance of this work, and to the National Science Foundation for a grant contributing to the support of this research during the academic year 1956–1957.

[1] For a concise statement of the meaning of a gravitational field independent of a gravitational source body, see: Einstein, Albert, *The meaning of relativity*, fifth ed., 140, Princeton Univ. Press, 1955.

[2] Callaway, J., Mach's principle and unified field theory, *Phys. Rev.* **96**: 778–780, 1954.

[3] Grunbaum, A., The philosophical retention of absolute space in Einstein's general theory of relativity, *Phil. Rev.* **66**: 525–534, 1957.

[4] Clifford, William Kingdon, *Common sense of the exact sciences*, 224–225, New York, D. Appleton, 1885.

[5] Jammer, Max, *Concepts of space*, 160–161, Harvard Univ. Press, 1954.

[6] We do not appeal to Clifford's philosophical or epistemological position or suppose that the following argument supports that position. Clifford is chosen to oppose Mach because his language, when dealing with the problem of space and matter, is suggestive of the view we have adopted.

mathematical premise of a theory of space-time and matter to be developed independently of dynamical concepts. We have noted that the general theory of relativity seems to conform to these requirements, and to this extent it makes Clifford's view plausible. However, if we are to insist on a display of the logical elements of the theory which contribute to either Mach's or Clifford's view, then something more substantial is required than the above observations. In this paper we will devise a method of formal evaluation of Clifford's ideas, and attempt to show how they are appropriate to general relativity and to a possible extension of the theory.

The philosophical issue between Mach and Clifford is a question of the scientific priority one should yield to the material particle, as compared with physical space-time. If we are to appeal to the logical structure of the relativity theory to provide evidence on behalf of Mach's or Clifford's position, then we must be willing to introduce a notion of logical priority which will serve to represent the intended philosophical priority. Our immediate objective is, therefore, to find a way of subjecting the general theory to an analysis which permits the formulation of a convincing criterion relative to which questions of priority among scientific constructs may be investigated. An axiomatization of the theory appears to be required, for an axiomatic framework will permit an examination of the logical structure of the theory,[7] while providing a natural meaning for relative priority among the terms or concepts it employs.

To treat a physical theory axiomatically, one must choose undefined primitives and axioms so as to define all of the remaining terms or quantities and demonstrate the relations which hold between those terms. In order that an axiomatization be helpful in discussing the question of priority, it must furthermore convey the intended status of each term relative to other terms. We hope to achieve the latter objective in the following way. *A term A will be said to have a status prior to a term B if term A is employed in the definition of term B.* According to this criterion, a primitive term is clearly prior to all other terms which require that primitive for their logical meaning. The writer feels that this criterion preserves the sense in which priority is generally understood; and at the same time permits one to subject physical constructs to an analysis in order to determine their priority in scientific theories.

A criterion of this kind is not sufficient, however, to guarantee uniqueness, for there may be a great many ways of axiomatizing a given physical theory. If two axiomatizations of the same theory employ significantly different primitive terms, then the question of priority would remain undecided so long as the above criterion is the sole consideration. This limitation need not threaten the value of the criterion, for we know in any case that a formal articulation cannot alone identify the conceptual content of a physical theory. A philosophical thesis which rests on a scientific priority should display its formal character as we have insisted, but in addition it must be persuasive on extralogical grounds. If the axiomatization of a physical theory is clearly artificial, or if it is unreasonably defiant of the understanding from which the theory has emerged, then we would find it difficult to accept, and would disqualify it as a representative of that theory.

A great deal may be said about the specific nature of extralogical requirements. Margenau[8] lists simplicity and elegance, multiple connection and extensibility among the metaphysical conditions which a physicist expects to be fulfilled by a satisfactory theory of nature. These requirements certainly should be reflected in an axiomatization which presumes to represent a physical theory. It is not difficult to appreciate how simplicity, elegance, and multiple connection fall into an axiomatic program, but the requirement of extensibility deserves special comment. We desire that an axiomatic examination of the structure of a theory be sufficiently penetrating so as to be suggestive of further theoretical research. This requirement has a tendency to obstruct the fulfillment of one other requirement given above. We stated above that an axiomatization of a theory should not seriously conflict with the understanding from which the theory has emerged. A requirement of this kind is necessary, but it could easily be interpreted to mean that one need only attempt to reflect an established tradition of scientific thought. This certainly is not our intention, for we will later use the axiomatization for the purpose of find-

[7] McKinsey, J. C. C., and P. Suppes, Philosophy and the axiomatic foundations of physics, *Proc. XI^th Int'l. Congress of Phil.* **6**: 49–54, 1953.

[8] Margenau, Henry, *The nature of physical reality*, Chapter 5, New York, McGraw-Hill, 1950.

ing a possible point of departure from the relativity theory. Furthermore, established scientific thought is rarely so uniform or so consistent as to permit a rigorous axiomatic treatment. For these reasons it is desirable, even necessary, for the axiomatization to deviate from some aspects of accepted scientific ideas. An axiomatization need not, therefore, be a simple catalogue of known formulas and rules, but should seek a point of view which is fundamental and suggestive enough to provide clues to its own extension.

Although we hope to employ significantly the criterion of priority proposed above, we see that it will not function satisfactorily alone. In addition, we must impose extra-logical considerations of a complicated kind in order that the criterion may be effectively employed. Perhaps it is not clear how these extra-logical conditions are intended to work. Since they involve judgments applied in particular cases, it may be helpful to consider a few possible axiomatic approaches to the general theory which the writer feels should be disqualified.

On the evidence of the general theory, it is easy to be persuaded that the properties of space-time and matter should not be considered distinct from one another as they were in Newton's physics, but rather that one should be taken prior to and definitive of the other. The Machian thesis that space-time is a relation between material objects is consistent with this view. Clifford's thesis that matter is a warp in space also expresses the idea, but with the suggestion of a different priority. When considering the many possible ways of axiomatizing the general theory, we will, therefore, be aware of the need to express adequately the desired unification of the properties of space-time and matter, and will avoid a formulation of the theory which minimizes this understanding. For example, it is possible to begin an axiomatization by adopting a primitive topology $\mathbf{T}_1$ which is specified on a set U of points which is employed as the basis of a model of the physical universe. The physical space is then represented by the pair $(U, \mathbf{T}_1)$, and will be called a physical topological space, or a PT-space. We may elaborate on the specific nature of the topology $\mathbf{T}_1$ by requiring as an axiom that the PT-space be locally a four dimensional continuum, and that the connectivity is both locally and universally appropriate to our understanding of the actual universe.[9] A PT-space cannot alone represent all that we mean by physical space, for it does not embody a notion of distance or metric. However, it clearly does have important space-like properties. We will, therefore, avoid an axiomatization of the general theory which provides for the construction of a PT-space, and which independently attaches dynamical primitives (i.e., primitives representing mass, momentum, etc.) to the points in that space. An axiomatization of this kind cannot represent the unification of the concepts of matter and space-time which we intend, for it takes primitive matter and bluntly puts it into a space which has been independently constructed.

There is a further aspect of the unification idea which permits the elimination of another possible axiomatization of the theory. The several properties of space should not themselves be logically separate. Therefore, an axiomatization is unsatisfactory which employs a primitive topology to construct a PT-space, and which independently employs a primitive displacement function for the purpose of giving the space a metric. To represent the general theory in this manner would imply the logical independence of physical topology and metric, and would therefore fail to express the dependence which Einstein's theory suggests.

We certainly cannot exhaustively consider all of the ways of analyzing general relativity. But the two possibilities which have been rejected and the reasons which have been given for their inadequacy ought to provide the reader with some idea of the standards we have adopted for judging a proposed axiomatization. With these examples in mind, it is helpful to generalize our standards somewhat as they relate to the notion of unification. If the properties of space-time (i.e., metric and topology) and matter are to be unified in the axiomatic sense of this paper, then there must be a unifying concept from which the above properties derive their meaning. The unifying concept should then be primitive in an axiomatic representation of that theory, so that other vital notions might be defined in terms of it. Therefore, a definition, *not an axiom*, must relate metric, matter, and topology if, in the writer's judgment, a notion of unification is to be justified. One might suppose that this criterion could be ranked as logical rather than extra-logical because we

[9] Axioms of this kind are formulated by requiring a homeomorphism (i.e., a congruence, or isomorphism) of the space $(U, \mathbf{T}_1)$ with an algebraic model which has been constructed from the numbers on the real axis and their real axis topology.

have specified the condition with apparent formal exactness. But this would be somewhat misleading. One can always examine an axiomatization and determine the priority of term A relative to term B by using our logical criterion; but if we employ the above criterion of unification to a given axiomatization asking: "Are the essential properties of space-time and matter related by axiom or by definition?" the answer will depend on our meaning of "essential properties," and what constitutes their relationship to one another. The proposed unification criterion is therefore extra-logical and requires the separate evaluation of each axiomatic representation of a physical theory. Of course, the other extra-logical requirements stated above enter into an evaluation, but it will be simpler to forego further discussion of them until we have all the details of a completed axiomatization before us.

In the following axiomatization, our primitive or elementary terms are the geometric notions (1) point, and (2) displacement function, the relata and relation of space-time. The displacement is eventually to be correlated with the metric of general relativity which is given by the *square* of the invariant interval between events (points). We will find that it is possible on this basis to define a satisfactory physical topology and specify the restrictions which must hold on it, consistent with general relativity. The availability of the real axis will be assumed so that the displacement function may assign numbers between all pairs of the primitively given points. The methods used in this axiomatization are similar to those familiar from distance geometry (metric spaces) although we require different axioms.

Since we are employing geometrical primitives together with a criterion of priority which elevates primitive notions to a rank higher than all defined terms, it is clear that we are attempting to express the relativity theory in a manner favorable to Clifford's thesis. The attempt succeeds in the simple sense that we can eventually define the matter tensor T_{kl} together with all of the other relevant tensor quantities of general relativity. But there are three defective aspects of this axiomatization which the writer finds unavoidable. These are not logical defects as such but are disturbing features of the axiomatization from the point of view of our extra-logical expectations. Two of the three defects appear as artificialities arising perhaps from our geometrical approach to the theory. They may therefore encourage one to believe that it should be possible to axiomatize general relativity in a natural way using matter-like primitives alone. The writer is sceptical about the possibility of this; but if it could be done, then the thesis of Mach would seem to gain an advantage from the proposed criterion. The specific nature of these two difficulties will be examined more carefully in Part III The third defect arises from a conflict between the principle of relativity (as we will formulate it in Part III) and the theory of relativity itself. The writer does not believe that this difficulty can be simply resolved by an alternative logical formulation of the theory. This question will also be discussed more thoroughly in Part III.

One further general comment. The following axiomatization provides two interesting features which are foreign to Einstein's general theory: (1) a definite meaning for the distance between two points independent of geodesics, and (2) a meaning for time direction independent of material (thermodynamic) systems. It may seem a bit strained to associate the first of these features with a differential field theory, but the need for a finite metric will be explained shortly. The second feature giving a meaning for time direction is distinctly desirable. It is not introduced arbitrarily, but follows from our solution of the special problems raised by the first feature stated above. The writer has been encouraged by this result, because it is entirely consistent with the view here presented that an important property of time such as asymmetry be given a structural meaning independent of material systems, and a status equal with the other independent properties of the space-time continuum.

II. PRIMITIVES AND AXIOMS

We have accepted the primitives: (1) point and (2) displacement function. The displacement function will assign real numbers, positive, negative, or zero, to all pairs of points in the universe in a manner that is arbitrary apart from the requirements of the axiomatization. Given any two points **a** and **b** in the universal set U of points, the numerical displacement from **a** to **b** is denoted by **ab**. Our axiomatization will proceed on the basis of the indicated primitives and real number theory. We do not assume that a space structure exists prior to the adoption of axioms which appropriately relate the

points in the universal set U with the displacement values given to all pairs of points in U. In this section we will discuss our choice of primitives and axioms and indicate some important theorems. The complete axiomatization is given in the appendix.

The first novel feature (referred to in Part I) of this axiomatization is a definite meaning for the distance between two points independent of, or at least prior to, a notion of geodesic. This order of definition follows from our use of primitive displacements which are finite rather than differential, and which are specified between all pairs of points in the universe. Since the equations of physics including those of general relativity are written in differential form, our choice of a finite primitive displacement (i.e., the distance, or metric) may be puzzling. However, if differential forms such as ds/dx are employed primitively, then one must independently assume the existence of a coordinate space relative to which a differential is meaningful. A coordinate space is understood to be a PT-space in which at least one coordinate system is defined. If we do not permit ourselves the luxury of postulating a space of this kind apart from the chosen metric ΔS^2, it does not make sense to employ primitive metrical displacements which are required to be differential. In the absence of an understood coordinate space it follows, therefore, that we must rely on finite primitive metrical displacements.

There remains, however, the choice of either selecting displacement values between all pairs of points in the universe, or somehow restricting the assignment in order to preserve the idea of a local field theory. The latter alternative requires the recognition of a special locale of each point in the universe of points and is objectionable for this reason alone. In addition it only artificially preserves the idea of a local field theory because the use of a primitive finite displacement between at least one pair of points implies an essentially non-local theory. For these reasons the latter alternative is rejected, and we employ finite primitive displacements between all pairs of points in the universe.

Choosing the primitives of an axiomatization may be compared to selecting the independent variables when proposing a Lagrangian for some physical system. The reason for choosing a certain set of variables is quite often related to the ease with which these variables, together with the proposed Lagrangian, lead to the desired equations as a consequence of variation. Similarly, the primitives of an axiomatization are chosen with some thought to the ease with which they, together with the proposed axioms, lead to desired theorems. In our case, the special reasons which have been given for the selection of a finite primitive displacement give additional significance to the choice. We are led to this choice by an attempt to represent general relativity axiomatically so as to express a particular understanding of the unification of space, time, and matter. Hence the use of geometrical rather than matter type primitives, and the rejection of a coordinate space which has a status independent of the chosen metric.

It is clear that we could not proceed on the basis of a primitive topology rather than a primitive displacement; for our method requires that these two spacial properties should not be independent, and hence that one should be defined by reference to the other. Since a topology is not strong enough to determine a metric, the inverse procedure is adopted. That is, we will define a topology (called $\mathbf{T}_1$) which is the topology of the universe of points, and which is logically dependent on the primitive displacement assignments. This is the usual order of dependence in the theory of metric spaces,[10] but the usual way of establishing this dependence is not satisfactory for Minkowski Space as will be later explained. For the present it is sufficient to say that the topology $\mathbf{T}_1$ is a family of subsets of the universal set U of points, and that every member of $\mathbf{T}_1$ is called an open set of the PT-space $(U, \mathbf{T}_1)$. The members of $\mathbf{T}_1$ (i.e., the open sets) will be explicitly defined by reference to displacement values as soon as we have examined the character of the first six axioms. The definition that is given will probably not incite immediate recognition; for the open set acquires a familiarity only after axiomatic restrictions have been formulated so as to make its properties transparent. We might say, however, that the word "open" does reflect the character of the open set as it will appear in the completed axiomatization. The word is used here in the sense that one speaks of an open interval on the real axis. Roughly, an open point set is one which does not include any of its boundary points. This will not serve as an adequate definition of an open set, but it

[10] Kelley, John L., *General topology*, 118–119, New York, D. Van Nostrand, 1955.

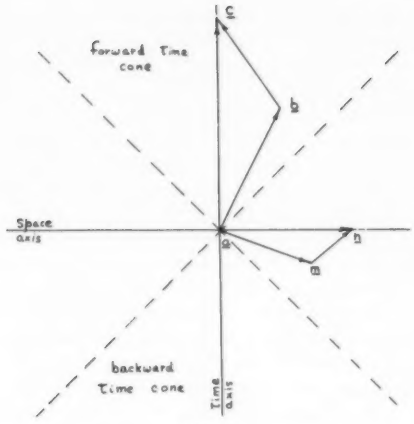


FIG. 1.

does convey part of the intended meaning. The topology T_1 is the family of *all* such open sets.

A neighborhood of some point $\mathbf{a} \epsilon U$ is any set $N(a)$ which includes an open set V of T_1, such that V contains the point **a** as a member.[11] Again in the context of the completed axiomatization the properties attributed to a "neighborhood" are such that this set adequately represents the familiar notion which that word suggests.[12] For instance, the first axiom provides for an intuitively understandable property of the space called the Hausdorff requirement. A space is called Hausdorff[13] if and only if for every two distinct points **a** and **b**, there exist disjoint neighborhoods of **a** and **b**. That is, there exists a neighborhood $N(a)$ of point **a** and a neighborhood $N(b)$ of point **b** such that $N(a)$ and $N(b)$ do not overlap.

Axiom 1: The space (U, T_1) is Hausdorff.

The second axiom will require no explanation. It states that the absolute magnitude of the displacement from any point **a** to any point **b** is equal to the absolute magnitude of the displacement from **b** to **a**.

Axiom 2: $|\mathbf{ab}| = |\mathbf{ba}|$.

Justification for axiom 3 will now be given in some detail; for it is not obvious that it is a justifiable proposition. In this discussion we will frequently refer to geodesic paths, the Minkowski Space of special relativity, and a notion of Minkowski Space extended to general relativity. The reference to complete space structures is given only to clarify the intention of axiom 3.

In ordinary metric spaces an axiom called the Triangle Law (or inequality) is assumed to hold. This inequality states that for all points **a**, **b**, and **c**: $d(\mathbf{ab}) + d(\mathbf{bc}) \geqslant d(\mathbf{ac})$, where $d(\mathbf{ab})$ is the linear distance ΔS between points **a** and **b**. Unfortunately, this rule does not hold for all triplets in the Minkowski Space of special relativity. To show this, consider the displacements from any point **a** to points in the corresponding positive region of Minkowski Space (where the signature is: $\Delta S^2 = \Delta x^2 + \Delta y^2 + \Delta z^2 - \Delta t^2$). If the triangle of points **a**, **m**, and **n** in figure 1 are assumed to lie in the x–y plane (perpendicular to the page), then the triangle inequality $d(\mathbf{am}) + d(\mathbf{mn}) \geqslant d(\mathbf{an})$ does hold; for in Minkowski Space this plane is Euclidian. But if the point **m** is chosen close to the forward or reverse time cone, then the sum $d(\mathbf{am}) + d(\mathbf{mn})$ might be less than $d(\mathbf{an})$. In the limit as point **m** approaches the intersection of the reverse time cones of point **a** and **n**, the sum $d(\mathbf{am}) + d(\mathbf{mn})$ approaches zero. That is, if the point **m** represents the emission of a light signal which subsequently travels to points **a** and **n** in Minkowski Space, then it is entirely possible that the invariant distance $d(\mathbf{an})$ will be positive, while $d(\mathbf{am}) = d(\mathbf{mn}) = \Delta S(\text{light}) = 0$. The triangle inequality $d(\mathbf{am}) + d(\mathbf{mn}) \geqslant d(\mathbf{an})$ is clearly not valid for this case.

Nonetheless, a triangle inequality does hold in a regular way for points in the forward time cone since that region is contained in a symmetric cone about a single dimension (i.e., there are no accompanying Euclidian planes). We will express the inequality in terms of the displacement adopted for this axiomatization rather than a linear distance, although the following argument would be valid in either case.

If point **b** of figure 1 is in the forward time cone of point **a**, and if point **c** is in the forward cone of point **b**, then $\mathbf{ab} + \mathbf{bc} \geqslant \mathbf{ac}$. The displacements **ab** and **bc** are, of course, negative, and **ac** is a number more negative than **ab** + **bc**.

This inequality may be understood by recalling that the displacement **ac** represents minus the square of the integrated time recorded on a clock moving along a geodesic from point **a** to point **c**.

[11] The set theoretical symbol "ϵ" means: "member of." That is, point **a** is a member of the universal set U.

[12] One precaution is required. A neighborhood of a point **a** is not necessarily a *small* neighborhood. It may include the whole universe of points. The universe is, in fact, a neighborhood of each of its points.

[13] *Ibid*, 67.

If a second clock were to travel from point **a** to point **b** along a geodesic and there be accelerated[14] so as to travel along another geodesic from point **b** to point **c**, then the second clock would record less time between events **a** and **c** than the first clock. This follows from considerations in special relativity, and is valid for any combination of paths the second clock may follow. Squaring the total time does not alter the inequality which expresses a simple physical property of clocks which are displaced along geodesics in Minkowski Space. We want to say that the correct displacement **ac** between points **a** and **c** is given by the clock which follows a single geodesic from point **a** to point **c** in this physical picture. From the point of view of our formalism, however, we will only say that the triangle inequality holds, and not presuppose that we have defined a clock. The physical analogy is only a guide at this point and will be further helpful when considering the possibility of extending the formalism to include a general Minkowski Space (i.e., the space of general relativity).

It appears that the inequality law might get us into trouble if we apply it without stipulating that all of the displacements must be forward (or all backward) in the time cone. That is, we do not mean to say that: $\mathbf{ac} + \mathbf{cb} \geqslant \mathbf{ab}$ where **ac** and **cb** are negative, and have the directional meaning given to them above. To show that a triangle inequality law does not lead to an ambiguity of this kind, we will assume that it holds for all negative displacement and demonstrate that the supposed difficulty does not exist.

Postulate: T.I. (triangle inequality)

For all points **a**, **b**, **c** if $\mathbf{ab} < 0$ and $\mathbf{bc} < 0$, then $\mathbf{ac} \leqslant \mathbf{ab} + \mathbf{bc}$.

When the axiomatization is complete we will be able to show that the displacement from any point to itself is zero. But for our present purpose it will be sufficient to demonstrate that the interval from any point to itself is non-negative. Consider any point **a** and suppose the displacement from point **a** to itself is negative. If we allow the three symbols **a**, **a′** and **a″** to represent the same point, then certainly $\mathbf{aa'} < 0$, and $\mathbf{a'a''} < 0$ as diagramed in figure 2. It

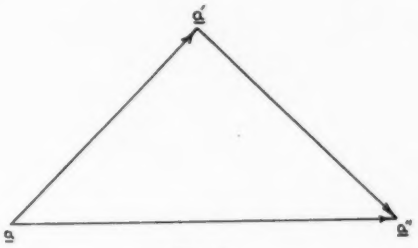


FIG. 2.

follows from T.I. that: $\mathbf{aa''} \leqslant \mathbf{aa'} + \mathbf{a'a''}$. In order to be consistent we would, therefore, have to deny that: $\mathbf{aa'} = \mathbf{a'a''} = \mathbf{aa''}$, which is contrary to our expectation that the displacement assignments are single valued.[15] Therefore, it is concluded that the displacement from any point to itself is non-negative.

Now consider any two points **a** and **b** such that: $\mathbf{ab} < 0$ and $\mathbf{ba} < 0$. The situation is characterized in figure 3 where **a** and **a′** are again the same point. Applying T.I. we get: $\mathbf{aa'} \leqslant \mathbf{ab} + \mathbf{ba'}$. That is, if **ab** and **ba** (or **ba′**) are negative, then **aa′** must also be negative contrary to our previous conclusion. *It follows that if any interval* **ab** *is negative, then the interval* **ba** *must be non-negative*, and hence an asymmetry is implied in a displacement which lies along the time cone.

T.I. may, therefore, be accepted without further stipulating that it concerns a sequence of displacements which proceed in one time direction. The postulate gives a meaning for time direction, and contains its own guarantee that it will apply in only one direction. For any two points **a** and **b** where $\mathbf{ab} < 0$, we may, therefore, distinguish point **b** as a successor of point **a**. Then, if any point **c** is known to be a predecessor of point **b** (i.e.: $\mathbf{cb} < 0$), no conclusion follows

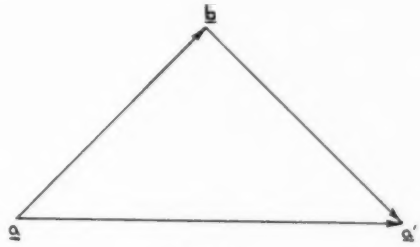


FIG. 3.

[14] We assume that the acceleration is instantaneous and has no effect on the reading of the clock. Or if it seems more satisfactory, let some other clock relay the time from the point **b** to point **c**. See Grübaum, A., The clock paradox in special relativity, *Phil. Sci.* **21**: 249–253, 1954.

[15] This expectation will be specifically formulated and adopted in axiom 5.

from T.I. as to the magnitude or sign of **ac**, since the displacement **bc** must be non-negative. But if point **c** is a successor of point **b**, then T.I. applies, and we shall know that point **c** is also a successor of point **a** with appropriate limitations on the magnitude of the displacement **ac**.

Reichenbach[16] has associated the idea of time direction with causality. The writer does not feel that causality as a philosophical thesis should be so directly concerned with the predecessor-successor relationship of temporal events; but there is some advantage to characterizing the above implications of T.I. in terms of cause or influence. If event **a** influences (or could in principle influence) event **b**, and if event **b** influences event **c**, then it follows that event **a** influences event **c**. On the other hand, if we know simply that events **a** and **c** both influence event **b**, than the causal relation of event **a** to event **c** is not determined. The use of this terminology is not harmful, but the reader is reminded that T.I. was suggested to us by the behavior of clocks following geodesics in Minkowski Space and not by appropriating the prior notion of "cause."

There is one difficulty with T.I. which arises from the fact that it is an inequality which applies on only one way around a given triangular circuit of points. This may be understood by taking a point **f** along the intended geodesic path **ac** in figure 1, and applying T.I. to displacements **af** and **fc**. If (in our axiomatization) **af** and **fc** are known, then the magnitude of **ac** is not uniquely determined by T.I. We certainly intend that **ac** should equal $(\sqrt{\mathbf{ab}} + \sqrt{\mathbf{fc}})^2$ in this case, but we have no guarantee that it will. There are ways of remedying this ambiguity by a simple strengthening of T.I.; or preferably, by the use of a modified axiom together with the invariant interval as a primitive displacement rather than its square. Our reason for adopting the weak form and permitting the ambiguity to remain will be discussed in Part III. In its present (weak) form, T.I. is valid and valuable for our purpose.

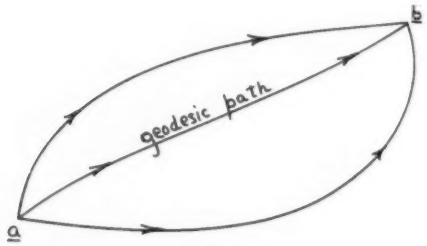


FIG. 4.

[16] Reichenbach, Hans, *The direction of time*, Berkeley and Los Angeles, Univ. of California Press, 1956.

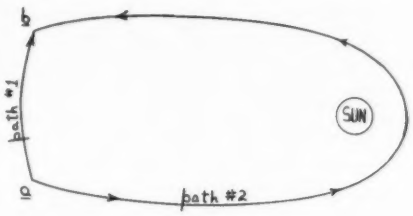


FIG. 5.

We will now entertain the possibility of extending T.I. to a general Minkowski Space. A general Minkowski Space will refer to the metric space-time of general relativity in which geodesics are defined by variational methods. Again, we appeal to the usual understanding of this space in order to evaluate the adequacy of T.I. The following argument is very general, but the writer feels that it is sufficiently convincing to justify the proposed extension. In considering this argument, it may be helpful to suppose that we have adopted the stronger form of T.I. If the reader is convinced on these grounds, then accepting an extension of the weaker form of T.I. should offer no difficulty.

First, consider a region such that there exists a unique geodesic path from point **a** to point **b**. We want to insure that T.I. leads to an assignment of the displacement **ab** which is consistent with the usual geodesic definition. In figure 4, three possible paths from point **a** to **b** are represented. From T.I. we expect that the displacement from point **a** to any point along either path would be equal to or less than the sum of an indefinitely large number of incremental displacements along that path. Since that sum would be the largest in absolute magnitude along the geodesic path from point **a** to **b**, the situation is the same as the one encountered in Minkowski Space. A clock following the geodesic path would record a longer time interval between events **a** and **b** than a second clock constrained to follow some other path. T.I. therefore applies in this case.

In general Minkowski Space there may not be a unique geodesic between two points **a** and **b**. Our interest in this case is to insure that T.I.

reflects the most reasonable choice that could be made among the alternatives presented.

The two paths represented in figure 5 are geodesics from point **a** to **b**. At this stage of generality it is assumed that path #1 is confined to a relatively flat space. We now maintain that if there is to exist another geodesic from point **a** to **b**, then this must be accomplished in general relativity by a path which encounters matter (such as path #2 in figure 5), and that a clock (clock #2) traveling over a path of this kind will record a shorter interval between events **a** and **b** than a clock (clock #1) traversing path #1. The differential expression for the proper time dt_2 of a clock on path #2 relative to the proper time dt_1 of a clock on path #1 is given by

$$dt_2 = dt_1 \sqrt{\left[\left(1 + \frac{2x}{c^2}\right)^{\frac{1}{2}} - \frac{\gamma_i u^i}{c}\right]^2 - \frac{u^2}{c^2}}$$

where: x and γ_i are gravitational and vector (dynamical) potentials, $u^2 = \sum_i u_i^2$, and the u_i's are the spacial components of the velocity of clock #2 relative to the chosen coordinate system.[17] If we adopt a spherical coordinate system with its origin at the center of the sun; and if, furthermore, the distribution of matter about the origin is symmetrical, then the Schwarzschild solution to Einstein's field equation may be consulted. In that coordinate system, the dynamical potentials are zero and the differential form reduced to

$$dt_2 = dt_1 \sqrt{1 + \frac{2x}{c^2} - \frac{u^2}{c^2}}.$$

Therefore, so long as the gravitational potential x is negative over the entire path of clock #2, our expectations will be fulfilled. That is:

$$\left|\int_a^b dt_2\right| \leqslant \left|\int_a^b dt_1\right|.$$

In the limit of large distances from the sun the gravitational potential comes to $x = -\frac{mc^2}{r}$ (where r is the radial distance from the sun). The sun is, therefore, a potential well in which clocks generally run slow relative to a clock which is stationed outside of the field. Does this field ever rise again near the origin to the extent of becoming positive, and possibly overcoming the terms in the above expression? A study of the solutions of the field equation suggest that it would not. Of course, the general theory is not satisfactory in its present form, and a complete theory including elementary particle fields may change the picture. However, we do expect that for any possible case, the value of the integrated path of any geodesic around the sun will be less in absolute magnitude than the time recorded on a clock outside the field.

In the case that there is a non-symmetrical distribution of matter in the region penetrated by clock #2 (two suns for example), then we must abandon mathematical solutions, and speculate on the basis of our results to this point. If matter in the form of one symmetric sun always has the effect of slowing down clocks which approach it, then we may feel reasonably safe in generalizing this effect of matter of clocks. This speculation is further supported by solutions of the field equation which feature continuous distributions of matter such as Einstein's cosmological model.[18] To the extent that this conjecture is justified, then T.I. will always settle the question of the non-uniqueness of geodesic paths in favor of a clock such as clock #1 in figure 5.

We have now considered two cases in general Minkowski Space. (1) Clocks constrained to move in some non-geodesic path, and (2) clocks which are drastically deviated along geodesic lines that penetrate a region of concentrated matter. What happens if we cannot make special assumptions about the field in the region of path #1? In that case, we are not guided by a preference about the particular selection for the value of the displacement **ab** which should be made by application of T.I. If there exists a longest time interval recorded among the class of all clocks leaving event **a** and returning via geodesics to event **b**, or if there exists a limit value of all possible geodesic intervals taken as positive numbers, then T.I. will serve to dis-

[17] Møller, Christian, *The theory of relativity*, 247, Oxford Univ. Press, 1952.

[18] From the signature of the Einstein universe, the relation between the time recorded on a stationary clock #1 and a clock #2 moving relative to the ponderable matter of the universe is:

$$dt_2 = dt_1 \sqrt{1 - \frac{v_r^2/c^2}{1 - r^2/R^2} - \frac{r^2 v_\theta^2}{c^2} - \sin^2\theta \frac{r^2 v_\varphi^2}{c^2}}$$

where: v_r, v_θ, v_φ are the coordinate velocities of clock #2, and R is the radius of the universe. Since the entire universe is confined within a radius R of clock #1 there is no case in which $dt_2 > dt_1$. See Tolman, Richard C., *Relativity, thermodynamics and cosmology*, 337, Oxford Univ. Press, 1934.

criminate against all of those values less than the largest, or less than the limiting value. That there will always be a largest, or a limiting value is guaranteed if we forbid positive potential singularities from appearing in the field. We are, therefore, sure that the strong form of T.I. will provide a unique selection for the displacement **ab**. It has been noticed that the weaker form removes the uniqueness that we have been working so hard to achieve. But this ambiguity goes (valuewise) in a different direction than the disqualified geodesics. It has nothing to do with the plurality of geodesics and will be discussed later in another connection.

From an examination of the above most general case it appears that our acceptance of T.I. is a declaration of a convention as to how we are to restrict the values assigned to negative displacements. This is true in a sense but not in the unrestricted sense of the word "convention," for not any rule will serve equally well. We have argued that this particular rule is appropriate to the axiomatization we are undertaking. It is especially suited to our attempt to express the general theory in a way that exhibits a special understanding; and we have been careful to insure that the metrical limitations it imposes on general Minkowski Space go over into the usual metric as we approach the limit of flat space. On these grounds, therefore, we expect the triangle inequality (T.I.) to hold for all negative displacements in the universe.

If proposition T.I. were accepted as an axiom directly, we would find it necessary to stipulate an additional condition on zero displacement assignments in order to confine these assignments to the boundaries of time zones. There is no obvious way to avoid the use of two separate axioms as a means of a requiring both T.I. and a zero assignment restriction. However, we prefer to employ a pair of axioms (axioms 3 and 4) which treat the zero displacements as though they too possess directed properties, thereby permitting them to be used together with negative displacements in a modified triangle inequality. This procedure also simplifies some of the theorems. Of course, it is not possible to associate a notion of displacement asymmetry too closely to that of numerical displacement values, for if we did it would be impossible to give a direction to zero assignments so long as axiom 2 is in effect. We therefore define a *successor directed* displacement by reference to the notion of limit point.[19] If we call the set of all successors of point **a** the successor set $S(a)$ of **a**, then any displacement **ab** is called successor directed if and only if point **b** is a limit point of the successor set of point **a**. The successor directed displacement will be designated $\overrightarrow{\mathbf{ab}}$. The reader should note that the symbol $\overrightarrow{\mathbf{ab}}$, is an abbreviation for the proposition that **b** is a limit point of $S(a)$; whereas **ab** is a number equal to the displacement from **a** to **b**.

We show in theorem 1 in the appendix that any negative displacement is successor directed as one would expect. That is $\mathbf{ab} < 0 \supset \overrightarrow{\mathbf{ab}}$. It follows that the proposition T.I. investigated above is a special case of the following adopted axiom.

Axiom 3: For all points **a**, **b**, and **c**, if $\overrightarrow{\mathbf{ab}}$ and $\overrightarrow{\mathbf{bc}}$, then $\mathbf{ac} \leqslant \mathbf{ab} + \mathbf{bc}$.

The advantage of this modified form of the triangle inequality follows only in conjunction with axiom 4 which requires that zero displacements are directed. We will call a displacement *directed* if and only if either $\overrightarrow{\mathbf{ab}}$ or $\overrightarrow{\mathbf{ba}}$.

Axiom 4: For all points **a** and **b** such that $\mathbf{ab} = 0$, the displacement **ab** is directed.

Our way of formulating the above pair of axioms is such that axiom 4 does not appear *ad hoc*. Actually, its major function is to confine zero displacements to the boundary of the time zone as explained above. In addition, axiom 4 will facilitate the proofs of a number of other theorems; but since the axiom does not seem to have a very fundamental status, it may be considered somewhat *ad hoc*.

The next axiom (number 5) provides for continuity of displacement assignments. Continuity is defined relative to the Topology $\mathbf{T}_1$, and the axiom is formulated by reference to two functions which are specified relative to any point **a**. The first function $d_a^{(1)}(u)$ is a function defined on the universal set U, such that the value $d_a^{(1)}(u)$ of the function at point $\mathbf{u} \epsilon U$ is equal to **au**. That is, for all **u**, $d_a^{(1)}(u) = \mathbf{au}$. The second function $d_a^{(2)}(u)$ is also defined on the universal set U, such that the value $d_a^{(2)}(u)$

[19] A limit point is defined relative to the topology $\mathbf{T}_1$. A point **p** is a limit point of set P if and only if every neighborhood of **p** contains a member of set $P - \{\mathbf{p}\}$.

of the function at point $\mathbf{u} \epsilon U$ is equal to **ua**. That is, for all **u**, $d_a^{(2)}(u) = \mathbf{ua}$.

Axiom 5: For any point **a**, and all points $\mathbf{u} \epsilon U$, the relative distance functions $d_a^{(1)}(u)$ and $d_a^{(2)}(u)$ are continuous and single valued.

It has been noticed that if a displacement **ab** is negative, then axiom 3 requires that **ba** is positive. We will call any displacement **ab** such that $\mathbf{ab} < 0$ or $\mathbf{ba} < 0$ a timelike displacement, and all other displacements spacelike. A zero displacement is spacelike, although it is distinguished among spacelike displacements by being directed. All timelike displacements are directed.

To say that a set A of elements is homeomorphic with a set B of elements is to require a topological congruence or isomorphism. Sets A and B are homeomorphic iff:[20] (1) there exists a function f which matches the elements of the set A in one-to-one fashion to the elements of set B, and (2) function f and its inverse f^{-1} are continuous. In the following paragraph, we will also have occasion to refer to a special set X which is defined to be the closed interval from 0 to 1 on the real axis.

We now define an important set called a chain. A set $Q \subset U$ is called a chain iff: (1) the displacement between any two points in Q is timelike, and (2) Q is homeomorphic with the set X.

Intuitively, a chain is intended to be a one-dimensional line of points which is directed into the time zone at every point. We will designate a chain by $C(ab)$, where **a** and **b** are the members of Q which correspond to the numbers 0 and 1 under the homeomorphism f. In this designation it is understood that $\mathbf{ab} < 0$ rather than $\mathbf{ba} < 0$. Some of the initial theorems in the appendix are devoted to showing that the points **a** and **b** of any chain $C(ab)$ are the end points of the predecessor-successor relationship, and that all neighborhoods about a point in a chain (other than end points) contain a predecessor and a successor also in the chain. Other expected properties of this important set follow from our definition.

When a chain $C(ab)$ is said to exist between points **a** and **b**, we do not mean to say that there is only one path from points **a** to **b** having chain properties. Indeed, there will be in general an infinite number of different chains from **a** to **b**, and each will correspond to what might normally be called the course of some observer whose route through Minkowski Space coincides with events **a** and **b**. When it is necessary to distinguish one chain route from another, we will use subscripts $C_i(ab)$ in the designation of the chain. At this point, not very much can be said about the alternative chain routes between two time-displaced points, but we must require the existence of at least one such chain.

Axiom 6: For all points **a** and **b** such that $\mathbf{ab} < 0$, there exists at least one chain $C(ab)$.

We have now presented the primitives and axioms 1–6. There are a total of eleven axioms, but it is important to pause here for the purpose of defining an open set in terms of our primitive displacement values and clarifying the meaning of the Topology $\mathfrak{T}_1$. We will also define other important concepts which are used for the purpose of proving interesting theorems, and summarize those theorems which can be established on the basis of these six axioms.

In order that a family of subsets be a topology $\mathfrak{T}$ for some space, it is important to insure that: (1) the union of any collection of members of $\mathfrak{T}$ is itself a member of $\mathfrak{T}$, and (2) that the intersection of any finite collection of members of $\mathfrak{T}$ is a member of $\mathfrak{T}$. We must be able to say that any union of open sets is an open set, and that any finite intersection of open sets is an open set. To provide for this feature, one might first specify a special family of sets called basic sets, and construct the topology by defining an open set as one which can be expressed as the union of a finite intersection of basic sets. The resulting family of open sets is then a topology. A basic set is itself open, but the family of basic sets (which is called a subbase) is only a subfamily of the topology.[21]

In our case, a basic set is specified by reference to some pair of points in the universal set U. If the points chosen are **a** and **b**, then the basic set of **a** and **b** is defined to be:

> The set of all points **c** such that $\mathbf{ac} < 0$ and $\mathbf{cb} < 0$. Designate: $B(ab)$

With the help of axioms 1–6 we will attempt to gain some understanding of the properties of a basic set. First we note the identity of the proposition "$\mathbf{ab} < 0$" and the proposition "$B(ab)$ is non-null." That is, $\mathbf{ab} < 0 \equiv [B(ab)$ is non-null$]$. To show this, let $\mathbf{ab} < 0$ for some point

[20] The abbreviation "iff" means: if and only if.

[21] Kelley, J. L., *op. cit.*, 49.

pair **a** and **b**. Then from axiom 6 we know that there exists at least one chain $C(ab)$. In lemma 2 in the appendix, it is shown on the basis of axioms 1–6 that for any chain $C(ab)$, the set $C(ab) \bigcap B(ab)$[22] is non-null. Therefore, $B(ab)$ is non-null. Conversely, if we say initially that $B(ab)$ is non-null, then there must exist at least one point **c** such that $\mathbf{ac} < 0$ and $\mathbf{cb} < 0$. Since it can be shown independently that $\overrightarrow{\mathbf{ac}}$ and $\overrightarrow{\mathbf{cb}}$, in this case (see thm 1 in the appendix), it follows from axiom 3 that $\mathbf{ab} < 0$. The identity is therefore established. From this identity and the asymmetry of negative displacements which we have demonstrated in connection with axiom 3, it follows furthermore that if $B(ab)$ is non-null, then $B(ba)$ is null.

These observations should help to clarify some of the essential properties of a basic set. In the predecessor-successor terminology, it is correct to say that $B(ab)$ is the set of all points which are both successors of **a** and predecessors of **b**. Intuitively we expect the set of *all* successors of point **a** to contain all of the points in the forward time cone of **a**, and the set of *all* predecessors of **b** to comprise all those points in the reverse time cone of **b**. If point **b** is a successor of point **a** (i.e., if $\mathbf{ab} < 0$), then these cones must intersect, and the set $B(ab)$ will contain the points in the intersection. If point **b** is not a successor of point **a**, then the cones cannot intersect, and the set $B(ab)$ will be null. It should be noted that points **a** and **b** are not themselves members of the intersection $B(ab)$ since the displacement from a point to itself is not timelike.

Perhaps it is now clear why set $B(ab)$ will qualify as an open set in the familiar sense. We expect that the boundary points of this set will be points which are related by light signals to either point **a** or point **b**; for light signals are indeed on the boundary of the predecessor-successor relationship. However, a basic set $B(ab)$ does not contain points which are separated from **a** or **b** by light signals; since the displacement in this case is zero rather than timelike. Therefore, set $B(ab)$ does not contain its boundary points, and it is open in that sense. In addition, it follows from axiom 5 (continuity requirement) that all boundary points of any $B(ab)$ are zero displaced to either **a** or **b**.

Again we will urge that although each set $B(ab)$ is an open set, there are a great many other open sets; namely all those sets which can be equated to the union of any finite intersection of basic sets $B(ab)$. The topology $\mathbf{T}_1$ is again the family of all open sets. Recall furthermore that a neighborhood of some point **a** is any set $N(a)$ which includes an open set V as a subset such that $\mathbf{a} \epsilon V$.

Since the point topology $\mathbf{T}_1$ is constructed by reference to basic sets $B(ab)$, and since they in turn are specified by reference to displacement values, we may feel satisfied to have adequately characterized a logical unification of the principle properties of physical space. In this unification the metric is given priority over that of topology in accordance with the criterion of priority that has been proposed. This means only that when metrical assignments are specified (consistent with this axiomatization) for some particular universe, all questions of topology will be automatically settled.

It was stated earlier that the usual way of establishing the dependence of topology on a metric is not satisfactory for Minkowski Space. It is customary in the theory of metric spaces[10] to define an open set by naming all points **c** such that $|\mathbf{ac}| < \epsilon$, where **a** is any point, and ϵ is any positive real number. Open sets defined in this way may be thought of as open spheres which surround the points specified (i.e., point **a** in the above case). The difficulty with this definition for Minkowski Space lies in the fact that zero displacements are not confined to go between each point with itself, but are assigned between many different points as well. For every point **a**, there exist points **b** throughout the Minkowski universe such that $\mathbf{ab} = 0$. If the usual definition of open set were employed in this case, the resulting topology would be unmanageable.[23] On the other hand, we have been able to construct a reasonable topology by taking advantage of the asymmetry of time-like displacements in the same way that one takes advantage of the asymmetry "greater than" and "less than" when specifying an open interval on the real axis. On the grounds of this simplicity, and the possibility of direct construction from displacement assignments, the topology $\mathbf{T}_1$ appears to be the most satisfactory one for this axiomatization.

We shall now define a special set called a

[22] For any sets A and B, their intersection $A \bigcap B$ is a set which contains all those elements which are members of both A and B.

[23] It is not helpful to regard Minkowski Space as an instance of a pseudo-metric space; for the triangle inequality does not hold between all points as we have seen.

periphery which is valuable for the formulation of the metrical properties of general Minkowski Space. In connection with this definition, it is important to appreciate some further implications of axioms 1–6. It follows from these axioms that any point $\mathbf{q} \epsilon B(ac)$ which is not in some chain $C(ab)$ is a boundary point of the successor set $S(c)$ of one and only one point $\mathbf{c} \epsilon C(ab)$. Point $\mathbf{c}$ is then called the lower vertex of $\mathbf{q}$ in $C(ab)$. Recall that the successor set $S(c)$ of $\mathbf{c}$ is the set of all successors of point $\mathbf{c}$. Similarly, the set of all predecessors of a point $\mathbf{d}$ is called the predecessor set of $\mathbf{d}$. Designate: $P(d)$.[24] Axioms 1–6 also require that for any point $\mathbf{q} \epsilon B(ab)$, $\mathbf{q}$ is a boundary point of the predecessor set $P(d)$ of one and only one point $\mathbf{d} \epsilon C(ab)$. Point $\mathbf{d}$ is therefore called the upper vertex of $\mathbf{q}$ in $C(ab)$. The situation is shown schematically in figure 6, where the vertical leg of the right isosceles triangle $\mathbf{a}\theta\mathbf{b}$ represents the chain $C(ab)$, and the area of that triangle represents the set $\{B(ab) - C(ab)\}$. If for every point $\mathbf{c} \epsilon C(ab)$ we construct two right isosceles subtriangles such as $\mathbf{a}\alpha\mathbf{c}$ and $\mathbf{c}\gamma\mathbf{b}$ in figure 6, then for every two points $\mathbf{c}$, $\mathbf{d} \epsilon C(ab)$, there will be a single intersection of a subtriangle constructed from $\mathbf{c}$ and a subtriangle constructed from $\mathbf{d}$ (i.e.: the intersection of lines $\mathbf{c}\alpha$ and $\mathbf{d}\beta$ in figure 6). If we agree to place all points $\mathbf{q} \epsilon \{B(ab) - C(ab)\}$ at that location in triangle $\mathbf{a}\theta\mathbf{b}$ which marks the intersection constructed from the two vertices of $\mathbf{q}$, then it can be shown that for any point $\mathbf{c} \epsilon C(ab)$, a member of $\{B(ab) - C(ab)\}$ will be located in the subtriangle $\mathbf{a}\alpha\mathbf{c}$ iff it is a predecessor of point $\mathbf{c}$. Furthermore, a member of $\{B(ab) - C(ab)\}$ will be located in the subtriangle $\mathbf{c}\gamma\mathbf{b}$ iff it is a successor of point $\mathbf{c}$. The set of all points located at a single intersection $\mathbf{c}\alpha$, $\mathbf{d}\beta$ is called a periphery, and is designated $K(cd)$. Formally, we define a periphery $K(ab)$ to be:

The set of all points $\mathbf{q}$ such that $\mathbf{q}$ is boundary to $S(c)$ and $P(d)$.

It can be shown that every point in $K(cd)$ is in $\{B(ab) - C(ab)\}$ if $\mathbf{c}$ and $\mathbf{d}$ are in $C(ab)$.

In the appendix, axioms 1–6 are used to show that the triangle in figure 6 is a consistent model which may be used to catalog all of the points in the set $\{B(ab) - C(ab)\}$. Theorems are proposed which elaborate on the characteristics of the space properties as they pertain to this

[24] Basic sets might very well have been defined by references to predecessor and successor sets since for all $\mathbf{a}$ and $\mathbf{b}$, $B(ab) \equiv S(a) \bigcap P(b)$.

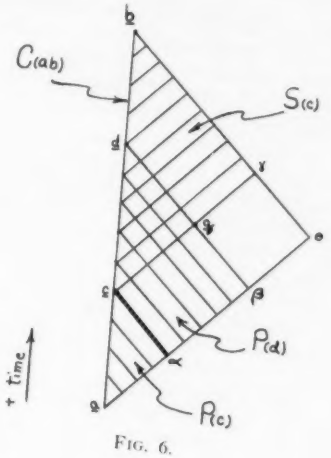


FIG. 6.

model, and particularly the decomposition of points into peripheral sets. For instance, it is shown that for all $\mathbf{q} \epsilon K(cd)$, $\mathbf{qc} = \mathbf{qd} = 0$, and that $\overrightarrow{\mathbf{cq}}$, and $\overrightarrow{\mathbf{qd}}$. Furthermore, it follows from axiom 3 that all successors of point $\mathbf{d}$ (in figure 6) are also successors of point $\mathbf{q}$; and that all predecessors of $\mathbf{c}$ are predecessors of point $\mathbf{q}$.

In order to get a firmer grasp of the geometrical concepts which are involved in the construction of a general Minkowski Space (especially the notions of a basic set, and a periphery set), it is helpful to consider the Minkowski Space of special relativity. In a special relativistic flat space the set of all events which are successors of some event $\mathbf{a}$ constitute a three-dimensional spherical volume, the boundary of which spreads outward from event $\mathbf{a}$ at the velocity of light. Among these events consider event $\mathbf{b}$ which is spacially coincident with event $\mathbf{a}$, but occurring at some later time. Now, the set of all predecessor events of $\mathbf{b}$ will constitute a three-dimensional spherical volume such that its boundary collapses on event $\mathbf{b}$ at the velocity of light. Therefore, the basic set $B(ab)$ will consist of a three-dimensional spherical volume which first spreads outward from event $\mathbf{a}$, reverses direction, and closes in on event $\mathbf{b}$. This picture will look very different from another Lorentz frame of reference; but since the topology is invariant under coordinate transformation, and since in special relativity we can always transform into a frame of reference which puts events $\mathbf{a}$ and $\mathbf{b}$ (where $\mathbf{ab} < 0$) at the same space point, our picture may be usefully extended to more general

situations. In this example, the periphery $K(ab)$ will be a two-dimensional, spherical surface embedded in a three-dimensional space. It is that surface which is defined by the coincidence of the surface of the outgoing volume with the surface of the incoming volume. Furthermore, if $C(ab)$ is taken to consist of the interval of events between **a** and **b** which are all spacially coincident with events **a** and **b**, then every spherical surface in $B(ab)$ which is symmetric about the spacial origin of **a** and **b** will be the periphery of some two events along $C(ab)$. From what has previously been said, it follows that the topological features of any periphery in flat-space are identical with the surface of a sphere in three-space. This observation is important to the sense of the eighth axiom: for in that axiom we state that all peripheries are spherical surfaces relative to some chain through each point $\mathbf{u} \epsilon U$.

We have already mentioned some propositions which follow from axioms 1–6 in connection with the apparatus needed to demonstrate the adequacy of the diagram in figure 6. This diagram is useful for the purpose of schematizing the decomposition of points in the set $\{B(ab) - C(ab)\}$ according to their membership in peripheries of $C(ab)$. Further results of axioms 1–6 are accountable, but we must first define three terms which have meaning only in the peripheral language: (1) the quadrifamily, (2) the horizontal peripheral line, and (3) the vertical peripheral line.

Peripheries may be grouped into families by designating all those pairs of points in some chain such that their accompanying peripheries contribute to the family. For instance, the four peripheries $M = K(c'd')$, $N = K(cd)$, $O = K(c'd)$, and $P = K(cd')$ in figure 7 determine the special family of peripheries in the shaded area of that figure. That family is called a *quadrifamily*, and is designated by $Q^i(cc', dd')$. Formally we say that periphery $K(xy)$ is a member of $Q^i(cc', dd')$ if both $\mathbf{x} \epsilon C_i(ab) \bigcap B(cc')$ and $\mathbf{y} \epsilon C_i(ab) \bigcap B(dd')$. The superscript i refers to the particular chain $C_i(ab)$ which includes points **c**, **c'**, **d**, and **d'**; for we expect that there will be more than one such chain, and for every such chain there is a characteristic family of peripheries. That is, the particular decomposition of points in set $C_i(ab) \bigcap B(ab)$ into a peripheral family depends on the choice of chain $C_i(ab)$. A unique quadrifamily is specified relative to a given chain $C_i(ab)$ for every four distinct points **c**, **c'**, **d**, and **d'** in $C_i(ab)$ such that $\mathbf{cc'} < 0$, and $\mathbf{dd'} < 0$.

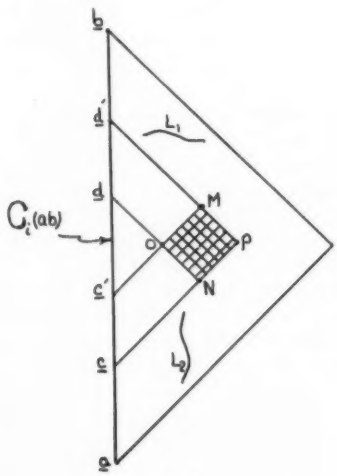


FIG. 7.

The class of all quadrifamilies relative to a given chain may be taken as the subbase for a topology defined on the accompanying periphery sets. This topology is called $\mathbf{T}_2^{(i)}$. An open family of $\mathbf{T}_2^{(i)}$ is any peripheral subfamily which is the union of finite intersections of quadrifamilies relative to $\mathbf{T}_2^{(i)}$. It is, therefore, meaningful to speak of the limit peripheries of a given subfamily of peripheries. The idea of a limit periphery of some peripheral subfamily corresponds roughly to that of a limit point of some region of the two dimensional triangular model used to represent the peripheral decomposition in figure 7. In that model the quadrifamily has the characteristics which one would expect of an open set in some topology.

Quadrifamilies are also used for the purpose of giving meaning to the proposition that one periphery is "above" or "to the left of" another periphery. In figure 7 for example, periphery M is above periphery N, and periphery O is to the left of periphery P. The formal definition of these terms are given in the appendix.

With these distinctions in mind, we can now define the special families L_1 and L_2 shown in figure 7. Family L_1 is called a *horizontal peripheral line*, and is defined by the following properties: (1) there are at least two peripheries in L_1, (2) for every two peripheries in L_1, one of them is to the left of the other, (3) for every two peripheries in L_1, there exists another periphery "between them" (this term is also defined in the

appendix), and (4) all limit peripheries of L_1 are members of L_1. The *vertical peripheral line* L_2 has similar properties, but in this case, the property of "above" replaces that of "to the left of." These definitions will facilitate a summary of the remaining interesting results of axioms 1–6.

It is possible to show that if two points $\mathbf{m}\epsilon M$ and $\mathbf{n}\epsilon N$ are such that $\mathbf{nm} < 0$ then M is above N, and every member of $B(nm)$ is also a member of a periphery in the accompanying quadri-family (see fig. 7). Furthermore, every two points taken from peripheries in a horizontal peripheral line have spacelike displacements. In case two points are taken from distinct peripheries in a horizontal family, then their displacement is positive spacelike. Although two points from the same periphery must have a spacelike displacement, it is not in general possible to prove that this displacement is positive. The latter result may seem contrary to intuition, but there are instances of non-isotropic geometry in which it is reasonable to expect two members of a single periphery $K(cd)$ to be connected by a zero displacement.

In addition, it is possible to prove, on the basis of axioms 1–6, that any chain $C_2(ab)$ different from chain $C_1(ab)$ will occupy a family of peripheries relative to $C_1(ab)$ which is a vertical peripheral line. That is, every point in $C_2(ab)$ will occupy one and only one periphery relative to $C_1(ab)$ and the family of all such peripheries will have the properties of a vertical peripheral line.

The results enumerated above have nothing to do with the dimension of the space (beyond the dimension of the chain) or the requirement of differentiability. Dimension and differentiability will be dealt with in the remaining axioms. Until now we have been concerned about the general properties of the model constructible on the basis of axioms 1–6 because it is important to be assured that a consistent and serviceable apparatus is available relative to which the more power axioms may be formulated. This assurance is particularly important for the axioms dealing with dimension; for we must state the requirement in two separate propositions, one of which (axiom 8) makes specific reference to the notion of a periphery. It will not be sufficient simply to require that general Minkowski Space is locally four-dimensional, because this does not inform us as to the relation of the metric to the property of dimension. For instance, we want to guarantee that the time zone represents only one dimension, and that the space zone includes three dimensions. We also want to insure that in some neighborhood about each point, all peripheries relative to some chain through the point are spherical surfaces. These features of Minkowski Space are usually assumed, and we may suppose that the assumption is supported by reference to the principle of equivalence. One argues first that the space is four dimensional, and then states that the differential line element at each point may be expressed by the signature $ds^2=dx^2+dy^2+dz^2-dt^2$ of special relativity. The difficulty with this formalization is that it separates the metric logically from topological features; and we have already stated our reasons for rejecting this kind of separation. It is far better to represent dimension as a condition holding on the metrical-topological model which has been constructed, and then (in conjunction with axioms guaranteeing differentiability) prove the existence of the particular line element shown above for some neighborhood about each point.

Axioms 7 and 8 provide for dimension. Not all of the terms and functions which need to be defined in order to state these axioms will be given in this section; for they are presented in detail in the appendix, and their reproduction here may only obscure our present purpose. Instead we will attempt to say what axioms 7 and 8 require without introducing more than a function h_i, and the notion of a two-dimensional spherical surface.

The function h_i is defined on the domain of points which are members of peripheries given relative to some chain, and over the range of those peripheries. That is $h_i(q) = K$ iff $\mathbf{q}\epsilon K$, where K is a periphery relative to a specified chain. The subscript i again indicates the particular chain route through Minkowski Space. In terms of the diagram in figure 6, the function h_i associates point $\mathbf{q}\epsilon K(cd)$ with the entire periphery $K(cd)$; and in figure 7, the periphery M is associated under h_i with every point $\mathbf{m}\epsilon M$. The function h_i will be single valued although not one-to-one. Axiom 7 then requires that: *both h_i and its inverse h_i^{-1} are continuous.* That is, open sets in $\mathbf{T}_1$ map into open sets in $\mathbf{T}_2^{(i)}$ under the function h_i and conversely, open sets in $\mathbf{T}_2^{(i)}$ map into open sets in $\mathbf{T}_1$ under the function h_i^{-1}. This means that if there exists at least one point in the periphery $K(cd)$ of some chain, then the universal set U is at least two

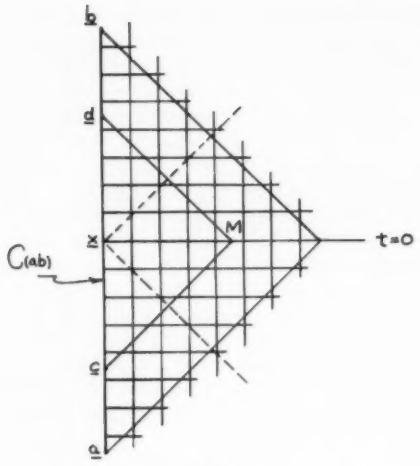


FIG. 8.

dimensional; for the open sets of $T_2^{(1)}$ are clearly two-dimensional regions by virtue of their defined relation to the one dimensional chain.

Axiom 8 employs an algebraic model of the two dimensional spherical surface embedded in a three dimensional Euclidean space. It requires that *for every chain, and for every member of the chain, there exists a subchain $C_i(ab)$ such that every periphery relative to $C_i(ab)$ is homeomorphic with the two-dimensional spherical surface.* We have already explained the sense of this requirement. It is suggested by a study of the topological features of peripheries in the flat space of special relativity, and it is needed to insure that peripheries do not become toroidal or plane in the limit of small neighborhoods about each point.

The above axioms are sufficient to guarantee that the space is locally four-dimensional, and possess the correct connectivity. With the aid of these axioms, it is possible to show that in some neighborhood about each point, all horizontal and vertical peripheral lines form three-dimensional subspaces of U. In the case of the horizontal line, we have already noted that all displaced assignments are spacelike. Therefore, a definite meaning can be given to the claim that the space which has been constructed has a three dimensional spacelike cross section over some finite region, plus a single time-like dimension. This claim is certainly less ambiguous when formulated relative to a metrical-topological structure of this kind, than it is when based entirely on the form of differential line elements. In the latter case it is difficult to see how this proposition would be stated in a satisfactory way, since it must rely on one's choice of world coordinates.[25] The advantage of the present formulation lies in the fact that it is stated in the invariant metrical language and refers directly to the physical topology T_1.

With the help of axioms 7 and 8, we can now demonstrate (by construction) the existence of a local Cartesian system of coordinates at each point such that one coordinate is confined to the time zone, and the remaining three coordinates are confined to the space zone. The construction proceeds in the following way: For some point $\mathbf{x}$, choose a chain $C(ab)$ containing $\mathbf{x}$, and points $\mathbf{c},\mathbf{d} \epsilon C(ab)$ such that $\mathbf{cd} < 0$. Points $\mathbf{c}$ and $\mathbf{d}$ will determine a periphery $M = K(cd)$ and we will say that all points in M have coordinate values r and t according to the following prescription: r (radius) $= \mathbf{dx} - \mathbf{cx}$; t (time) $= \mathbf{dx} + \mathbf{cx}$. In accordance with this scheme, the peripheral line such that $t = 0$ will identify a sequence of peripheries which approach $\mathbf{x}$ in the limit, but which never enter the time zone. These peripheries require two extra parameters to identify uniquely all of their points. We will, therefore, have three space coordinates and one time coordinate. This may be checked by noting that if $r = 0$, the resulting sets of points is $C(ab)$ which is a one dimensional set. To specify the remaining θ and φ coordinates, we use the homeomorphism provided by axiom 8. If it is required that the usual angular coordinates of the two dimensional spherical surface are carried over into the point set $\{B(ab) - C(ab)\}$ by the homeomorphism, then the full range of coordinates r, θ, φ, t will be satisfactorily specified. We need only then perform the usual transformation from r, θ, φ, t to Cartesian coordinates x, y, z, t in order to complete the Cartesian construction. It should be noted that the construction is Cartesian only in the sense that it provides a Cartesian way of classifying or identifying each point. It tells us nothing definite about the metric, so we do not have to be concerned about the fact that displacements used to define it are the square of the invariant interval. The system x, y, z, t so erected is called a *chain coordinate system* relative to $C_i(ab)$, and is denoted by $O_i^w(a)$ (where $\mathbf{a}$ is the origin

[25] Schrödinger, Erwin, *Space-time structure*, 85, Cambridge Univ. Press, 1950.

of the coordinates, and i is the chain which serves as the basis of construction and ω indicates the particular choice of coordinates θ and φ). If point **p** is a point identified by the coordinates of $O_i^{\omega}(a)$, then we will say for brevity: **p** id $O_i^{\omega}(a)$.

We now define a metric function S_a^2, relative to any point **a**, which is equal to $d_a{}^{(1)}(u)$ over the set of all non-predecessors of **a**, and equal to $d_a{}^{(2)}(u)$ over the set of all predecessors of point **a**. This means that S_a^2 will be equal to the displacement **au** for all points **u** except those points which are predecessors of **a**. In the latter case, $S_a^2 = -\mathbf{au} = \mathbf{ua}$.

A chain coordinate system $O_i^{\omega}(b)$ will be called *metrical* iff the function S_a^2 is differentiable relative to $O_i^{\omega}(b)$ for every **a** id $O_i^{\omega}(b)$.

Axiom 9: For every point **a**, there exists a chain including **a** and a choice ω of θ and φ coordinates such that $O_i^{\omega}(a)$ is metrical.

The ninth axiom does not state simply that the metric function is differentiable relative to some well-defined coordinate system, but states more generally that *there exists* a coordinate system of the form $O_i^{\omega}(a)$ relative to which S_a^2 is differentiable. This distinction is important to the sense of our discussion to follow in Part III. It is important to realize that there will be a great many chain coordinate systems with their origins at point **a**, and that it is only possible for a subclass of those systems to be differentiable relative to any given function.

From the existence of a metrical chain coordinate system $O_i^{\omega}(a)$, it follows that the function S_a^2 may be expressed in a Taylor expansion about point **a**. Suppose for some **a** we let the metrical chain coordinate system $O_i^{\omega}(a)$ also be denoted by $[klm]$ where $k, l, m = 1, 2, 3, 4$ (i.e., $x^1 = x$, $x^2 = y$, $x^3 = z$, $x^4 = t$). Then, since the displacement from **a** to itself is zero, and since S_a^2 is a locally even function in all coordinates, it follows that the first term appearing in the Taylor expansion will be second order. Therefore, the differential form of S_a^2 at point **a** may be expressed by the quadrature $dS_a^2 = g_{kl}dx^k dx^l$, where we use the summation convention over indices which are repeated in a product.

The array of terms g_{kl} defined above is called the *covariant metric tensor* specified at point **a**. It is shown in the appendix that the determinant $|g_{kl}| < 0$ for the coordinate system $[klm]$; and so it is possible to assure the regularity of the *contravariant metric tensor* g^{kl} which is equal to the cofactor of g_{kl} divided by $|g_{kl}|$.

The quantities g_{kl} and g^{kl} given above are clearly tensors as can be seen by a regular transformation of coordinates, and subsequent substitution of the differentials $dx^k = \dfrac{\partial x^k}{\partial x^r} dx^r$ into the invariant expression $dS^2 = g_{kl}dx^k dx^l$; but they are only so far defined at the point **a** relative to the metrical coordinate system $O_i^{\omega}(a) = [klm]$. In the appendix our definition is extended in such a way that the covariant metric tensors g_{kl} is specified at every point **p** id $O_i^{\omega}(a)$. We may then speak of the function g_{kl} over that subspace of points.

Axiom 10: For any metrical chain coordinate system $O_i^{\omega}(a) = [klm]$, the function g_{kl} is continuous and differentiable.

It is now possible to define the affinity $\Gamma_{kl}{}^m$, the contracted Riemann-Christoffel tensor R_{kl}, and finally the matter tensor T_{kl} at point **a** relative to the metrical system $O_i^{\omega}(a)$.

$$\Gamma_{kl}{}^m = \tfrac{1}{2}g^{mn}\left[\frac{\partial g_{kn}}{\partial x^l} + \frac{\partial g_{nl}}{\partial x^k} - \frac{\partial g_{kl}}{\partial x^n}\right]$$

$$R_{kl} = \Gamma_{kn}{}^m\Gamma_{ml}{}^n - \Gamma_{kl}{}^m\Gamma_{mn}{}^n + \frac{\partial}{\partial x^k}\Gamma_{lm}{}^m - \frac{\partial}{\partial x^m}\Gamma_{kl}{}^m$$

$$T_{kl} = \frac{1}{K}\left[\tfrac{1}{2}Rg_{kl} - R_{kl}\right]$$

where: $R = g^{kl}R_{kl}$

$$K = 2.073.10^{-48}\ \text{sec}^2/\text{gm cm}$$

The components of the tensor T_{kl} so defined at point **a** are the dynamical quantities stress, pressure, momentum, and mass-energy density of a material system at point **a**. If we say that T_{kl} is zero, then we mean that every component is zero; and if we say that T_{kl} is non-zero, then we mean that some (not necessarily all) of its components are positive or negative numbers. As the axiomatization stands, it would appear that we have accomplished our objective by successfully defining matter on the basis of a spacial construct of the universe. But the universe we have constructed admits many situations which we do not want to permit in the actual universe; for not all non-zero matter tensors can represent real matter as it is now understood. We will, therefore, distinguish real and fictitious matter, and propose an axiom

(axiom 11) which is calculated to eliminate the latter.

Axiom 11: For any metrical chain coordinate system $O_i^w(a) = [klm]$; if the matter tensor T_{kl} is non-zero, then T_{44} is positive.

The remainder of the axiomatization consists of the construction of the local inertial system at each point **a**, and a proof that the T_{44} component of the matter tensor is positive in the inertial system if it is positive in the metrical system $O_i^w(a)$ from which the inertial system is constructed. The properties characterizing a local inertial system (denoted $(\mu\nu\sigma]$) are: (1) $g_{11} = g_{22} = g_{33} = -g_{44} = 1$ and $g_{\mu\nu} = 0$ for all $\mu \neq \nu$, and (2) $\partial g_{\mu\nu}/\partial x^\sigma$ for all $\mu, \nu, \sigma = 1, 2, 3, 4$. Construction of the system $[\mu\nu\sigma]$ follows from $O_i^w(a) = [klm]$ via two transformations. The first transformation goes to a coordinate system such that the first derivative of the metric tensor is zero, and the second transformation diagonalizes the tensor to the form shown above.

Finally, it should be noted that we have not mentioned a general transformation group, or proposed axioms which serve only to qualify that group. Our interest has been to find those axioms which impose limitations on the primitive displacement assignments, consistent with our understanding of the invariant metric of a general relativistic universe. It is not important to those restrictions whether we do or do not entertain the broadest group of coordinate transformations permitted by a Riemannian calculus. It is necessary only to guarantee the regularity of those transformations which are needed to prove special theorems. In particular, it is important to show the existence of a local inertial system at each point, and the correctness of the sign of T_{44} in that system; for otherwise the axioms adopted would be inadequate to their task.

III. EVALUATION

An evaluation of the axiomatization will proceed in two directions: (*a*) an examination of those deficiencies in the axiomatization which appear to arise from a conflict between the general theory and our geometrical-topological approach to it, and (*b*) that deficiency which the writer believes is attributable to the failure of general theory to accomplish its objectives. The *ad hoc* axiom 4 is not here counted as a deficiency. It is certainly displeasing to have to add it for the sole purpose of guaranteeing that zero assignments are restricted to the boundary of the time zone, but the writer does not believe this to be an indication of fundamental difficulty. In the discussion which follows we will again refer to Clifford as an exponent of the view here represented, and set him opposed to the alternative we see in Mach.

Deficiencies under (*a*): If the Cliffordian view is correct, then it is a space property on which the axiomatic foundations of a theory of space-time and matter must rest. If we are correct in assessing metric and topology to be the essential space properties,[26] then only metric can serve as a useful primitive notion, for certainly a topology is not strong enough to define a metric. Furthermore, in the opening paragraph of Part II we explained the need to employ finite metrical displacements. This is because differentials have no meaning unless at least one coordinate system is given an equal status with the differential form. Aside from the undesirability of considering coordinate systems in this fundamental role, it is probable that an axiomatic program using differential primitives will lead either to an ambiguity, or the importation of a topology which is not logically dependent on metric, or a primitive meaning for the matter tensor. Therefore, Clifford's view leads rather directly to the use of finite primitive displacements. However, the general theory of relativity is contained in differential form.

This formal discordance leads to some technical difficulties which appear in our axiomatization in the following way. Aside from meeting the requirements of the axioms, the particular value of some displacement **ab** does not directly influence the component values of the metric or matter tensors at point **a**. The differential coefficients (i.e., tensors) have been defined by ultimate reference to the finite displacements, but any given displacement value can be changed without necessarily affecting the value of the differential form. For instance, suppose we have a coordinate system established with an origin at point **a**, and another point **b** is given on the x-axis (space-zone). Let $\mathbf{ab} = +10$. Then axiom 5 requires that $d_a{}^{(1)}(u)$ is continuous at **b**

[26] We specifically reject the coordinate system as an essential property of space. The "point" is taken for granted as a geometrical primitive which does not hinder the extension of unification to include different space properties. That is, we have not felt the need to require that the concept "point" be defined in terms of a distance function or *vice versa*.

so that points close to **b** will have displacements values from **a** close to + 10. Now the g_{11} component of the metric tensor is defined as the limit of the ratio of values of $d_a{}^{(1)}(u)$ to values of the x-coordinate as x approaches zero. It is consistent with the axiomatization to imagine a change of **ab** from + 10 to say + 11, such that the continuity requirement remains fulfilled, but that the differential coefficient g_{11} is unaffected. Since, furthermore, the example portrays a symmetry with respect to arguments **a** and **b**, the tensor components of g_{kl} at **b** may also be unaffected by a change. This lack of unique relationship between the finite displacement magnitudes and the differential coefficients is particularly evident in the space zone, although it is true to some extent in the time zone as well. Had we chosen the strong form of Postulate T.I. (thereby removing all ambiguity following from the triangle inequality), the specific magnitudes of negative displacements would have appeared to be unique; but we preferred to use the weak form of T.I. because our doing so would help to clarify the character of the ambiguity arising from our use of finite primitive displacements for the purpose of axiomatizing a differential field theory.

The second deficiency under (*a*) is related to the second novel feature that has been injected into general Minkowski Space by this axiomatization: time cone asymmetry. We found that an asymmetry followed directly from axiom 3 which was adopted for reasons unrelated to asymmetry as such. Furthermore, we found that this asymmetry led to a natural topological construction, and that it proved invaluable in many special theorems that had to be established. Yet again an inherent time direction is foreign to general relativity, so that we are led eventually to a technical difficulty which appears in the definition of the metric function $S_a{}^2$. To specify $S_a{}^2$ correctly, we had to take the precaution of reversing the sign in the predecessor time zone, thereby removing any functional dependence of the relativistic tensors on the property of negative displacement asymmetry. This is certainly an artificiality in our axiomatization; but it is not serious except as it draws attention to an opposing account of the status of a property of physical time. The relativity theory characterizes space-time as a four dimensional continuum wherein the space and time are given equal parametric status. The theory tends to emphasize those properties which space and time have in common, while neglecting their apparent differences. As a result, we are led to a statistical interpretation of time asymmetry contrary to the structural interpretation which characterizes the present treatment. We refer to the need to correct the function $S_a{}^2$ as a deficiency in the axiomatization only because our treatment is somewhat unnatural to the theory which is being axiomatized.

Apart from the reasons which may determine the choice between matter or structural defined time asymmetry, one might hope that an asymmetry can be significantly related to the formal body of physical theory, for it is just this kind of foundation that a pure statistical theory seems to lack. The logical formalism of statistics does not contain implicit asymmetries any more than does the formalism of particle mechanics. Generally, statistical definitions of time asymmetry appeal at a critical point to some aspect of empirically given time direction, thereby rendering a theoretical meaning non-existent.

Deficiency under (b): We now claim that axiom 9 is the focus of an extra-logical difficulty which arises from a limitation of the theory. Recall that axiom 9 provides for the existence of at least one chain coordinate system at each point relative to which the metric function is differentiable. It is not so important that the axiom mentions *one* coordinate system, but it is important that it distinguishes a proper subclass of the class of all chain coordinate systems.[27] To understand why this proposition conflicts with the aspirations of general relativity, we will briefly attempt a two-part formulation of the principle of relativity. This principle has two logically distinct features which are very frequently confused. They are: (1) particle priority, the older meaning of the idea of relativity, and (2) invariance, the modern meaning which has become prominent since Einstein.[28]

(1) *Particle Priority:* The problem of a referent for particle motion gave rise to a dispute in scientific philosophy. Newton solved the problem by introducing the notion of an absolute space and an absolute time relative to which particle motion is defined. Those who opposed the solution preferred to define the motion of a particle in terms of other particles. We will designate this latter proposition PPP, and refer

[27] It will be shown later that not all chain coordinate systems can be metrical.

[28] Margenau, H., and R. Mould, Relativity, an epistemological appraisal, *Phil. Sci.* **24**: 297–307, 1957.

to it as the principle of particle priority; for the issue of PPP is clearly one of the scientific priority which must hold between particles and space properties. It asks: should the properties of space be defined by reference to particles (Mach) or should particle motion be defined by reference to a prior absolute space (Newton)? To the extent that this issue is of current interest, it is not a question of observables as Mach would certainly have insisted. We have not yet mentioned anything in this paper which is strictly speaking an observable. The things given to us in experience are colors, odors, sounds, etc., and these are not at all like points, negative displacements, mass, or momentum. The latter are constructs employed by a physicist to formulate a concept of nature which may be projected by means of rules of correspondence to fit, order, or arrange his experiences thereby making them intelligible.[29] There is furthermore no rule of correspondence going from any single element in our experience to any single element in the formalism. The physicist cannot point to something and say "there is a T_{44} component of the matter tensor," as though it were a flower in the field. Scientific theories certainly do relate to experience, but only in a very complex way. They must be experimentally checked in context with the aid of a complete formalism relating the various constructs of the theory. It follows that if no isolated construct in the formalism is observable; that is, if we cannot go into a physics laboratory and see whether there really is a T_{44} matter tensor component there, then the question of scientific priority can have little to do with those constructs which are directly observable and those which are not. To be reasonable a notion of scientific priority must relate to some compelling aspect of the theory, for a physicist's concept of nature characteristically depends on the logic of construct relationships for a major part of its meaning. Therefore, we interpret the issue of PPP to involve the question of the logical priority which holds between particle properties and space properties on the level of axiomatic foundations.

(2) *Invariance:* The modern meaning of the principle of relativity is related to the extension of the laws of nature over the range of particulars. In the sense of an axiomatization it requires that the axioms apply equally to all individuals of the same kind. We will designate it GInv, and refer to it as the general principle of invariance; for it demands that physical laws are form invariant under substitution of different particulars which are judged to be the same in kind. This requirement does not mean very much by itself but applies when we have decided what are properly regarded as the same kinds of things. For instance, if one believes that our sun is not a unique object in nature, then according to GInv he would expect the laws of nature to apply with equal force to all suns, and the axiomatic propositions of a physics would therefore extend equally over the range of individual suns. Alternatively, we could say that the laws of nature are form invariant under transformation from one sun to another. This idea is certainly not new in physics, but it has been recently associated very closely with Einstein's principle of relativity owing to the latter's application of GInv to the class of all coordinate systems. In accordance with it, he requires that the laws of nature display indifference toward coordinate systems by invariant formulation with respect to transformations between coordinates.

What is Einstein's justification for believing that the class of all coordinate systems are the same kind of thing? To answer this question it will be helpful to elaborate on the distinction between metrical space and coordinate space. This distinction is important and rather straightforward. Consider an algebraic model of a two-dimensional Euclidian Space. In order to construct such a model, we first choose our elements (i.e., those things which function as points in the usual sense of a space) to be ordered pairs of numbers from the real axis. The total universe (of this space) is then the set of all ordered pairs (xy) such that $x \epsilon N$ and $y \epsilon N$, where N is the set of real numbers. We say that the space of ordered pairs is the Cartesian product of the real axis with itself and refer to it as the universe $U_2 = NXN = \{(x, y): x \epsilon N \text{ and } y \epsilon N\}$.[30] A product topology $\mathbf{T}_p$ may now be assigned to U_2 in the following way. Let V and W be any two open sets on the real axis. Then the set of ordered pairs $(v, w) \epsilon U_2$ with a first member in V and a second member in W will be an open set in $\mathbf{T}_p$. That is, if V and W are open sets in N,

[29] Margenau, H., *op. cit.*, chapters 4–6.

[30] In general, the Cartesian product of sets A and B is: $A \times B = \{(a, b): a \epsilon A \text{ and } b \epsilon B\}$. See Kelley, J. L., *op. cit.*, 7.

then $VXW = \{(v, w): u \epsilon V$ and $v \epsilon W\}$ is an open set in the topology T_p. It is easily shown that T_p is a topology.[31]

The above construction yields a topological space of elements (x, y) such that the space is homeomorphic with a two dimensional Euclidean plane. We will say, furthermore, that it is a *coordinate* space because the numbers x and y of each element (x, y) can serve as coordinates in the usual way. However, it is not yet meaningful to speak of the distance between two elements (x, y) and (x', y'). Hence, the space cannot be said to have a metric. In order to provide for a metric it is necessary to add that for every two elements (x, y) and (x', y'), there is number specified by a function $D = D[(x, y), (x', y')]$ called the distance between the elements. The function D is the distance function, and in the special case of a Euclidian geometry, $D = \sqrt{(x-x')^2 + (y-y')^2}$. Various types of non-Euclidean geometry follow from different choices of D although the function is not unrestricted because it should be required to be continuous relative to the topology T_p.

The above example provides the contrast between a coordinate space which does, and one which does not have a metric. Logically, the metric is a numerical relationship holding *between* two elements of the space, whereas the coordinates are numbers which serve only to distinguish one element from another.

In the present axiomatization the distinction between metric and coordinates certainly remains, but we have proceeded in a somewhat different way to construct our model of Minkowski Space. Our displacement function, which is analogous to the distance function in the above example, is taken as a primitive together with the primitive elements or points of the universe. Next, we defined a topology T_1 by reference to the displacement function; and only then did we identify points systematically by the use of a coordinate system. Prior to the construction of a coordinate frame, we identified points exclusively by letters **a**, **b**, **c**, etc. These two means of point identification (i.e., the Cartesian method and our non-systematic lettering method) are to be regarded as logically similar.

In accordance with this understanding of a coordinate system, it is reasonable to accept the Einstein view that all coordinate systems are to be treated as the same kind of thing; for certainly, if a coordinate system is only a convenient labeling of points, then any other labeling should offer no discriminating features to the physicist so long as he does properly distinguish points (i.e., properly distinguish means: correctly relate topologically to other points). As we have noted, this view together with GInv leads to an idea of the invariance of physical law under general coordinate transformation or general covariance.

We have now presented two aspects of the principle of relativity: PPP and GInv. They are logically distinct because the first deals with the question of priority, and the second has to do with the range or extension of the axioms over the individuals which are judged to be essentially the same in kind. Now what are the objectives of general relativity? Does the theory attempt to satisfy PPP, GInv, or both principles? The answer is that it attempts to satisfy both. This is done by identifying the objectionable absolute space of Newton with an absolute (i.e., high priority) coordinate space. The Einstein argument[32] is that a coordinate system is to be regarded as a convenient labeling of points, and therefore: (1) coordinate space cannot be physically significant or have a high priority status in the foundations of physical theory, and (2) the laws of nature must be insensitive to coordinate notation, or invariant under coordinate transformation. These two propositions result from the application of PPP and GInv to coordinate systems. They will be designated $\sim$CP and CInv, and referred to as the "denial of coordinate priority," and the "principle of coordinate invariance." It is to be noticed that whereas PPP affirms the high priority of particles, $\sim$CP only denies the high priority of coordinate systems. We therefore have four propositions which are roughly associated with the principle of relativity. They are listed below:

PPP: The principle of particle priority.

GInv: The general principle of invariant physical law.

$\sim$CP: The denial of coordinate priority
CInv: The principle of coordinate invariance.
} Einstein's Relativity

The writer believes that the identification of Newton's absolute space with a high priority co-

[31] *Ibid.*, 89.

[32] Einstein, A. and others, *The principle of relativity*, 117, New York, Dover, 1952.

ordinate system rather than a metric or topological space is historically justifiable, and certainly reflects the modern understanding. Contemporary physicists will generally agree that a coordinate system is only a convenient labeling of points and is not related to the basic stuff of the universe. The relativity theory is believed to furnish the evidence for that understanding. However, the evidence from general relativity is not very convincing as we will attempt to show in the forthcoming analysis. On the face of it, our representation of Clifford's views should not prejudice the case; for this view permits one to state a great many significant propositions in a metrical or topological language thereby delaying the introduction of coordinates. However, in order to define all of the quantities in the general theory which are intended to have good physical meaning, we had to violate the principle of relativity as formulated under both ~CP and CInv.

Violation of ~CP (denial of coordinate priority): We do not interpret proposition ~CP to mean that the physicist should in no case define a special class of coordinates for his convenience. It does mean, however, that any quantity which is defined relative to a special coordinate class, and which depends critically on some novel characteristic of the special class, cannot have fundamental physical importance. If we examine our axiomatization in the light of the adopted criterion of priority, then we find that ~CP has been violated; for the tensor quantities g_{kl} and T_{kl} are indeed defined by reference to a very limited coordinate class. It is assumed that tensors do rank higher in physical significance than ~CP would grant to a special class of coordinates. Yet it remains essential to the mathematical meaning of these terms that there exist a coordinate class with special characteristics relative to which they can be defined. We refer to the special differentiability characteristics of the metrical chain coordinate class. There can be no such mathematical property imposed on the widest possible coordinate class which includes a non-systematic lettering of points ($a, b, c \ldots$ etc.) as a member system. Clearly, tensors g_{kl} and T_{kl} cannot be recognizably defined in such a system. It is in this direct sense, therefore, that *~CP* has been violated in our axiomatization, and is necessarily violated by the Einstein relativity theory.

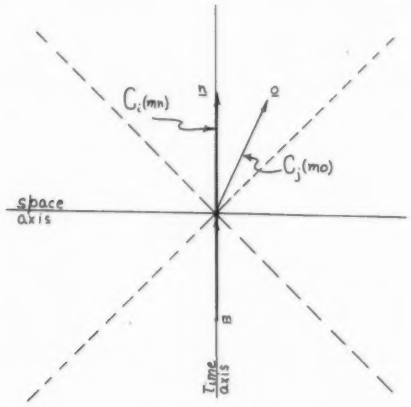


FIG. 9.

Violation of CInv (the principle of coordinate invariance): The analysis of proposition CInv will proceed along somewhat different lines than ~CP. All coordinate systems in the general coordinate class are given equal status. CInv then requires that the axioms must not discriminate among members of this class. Importance here is attached to the range of the axioms and not the intended priority status of the terms or quantities involved.

The most general class of coordinates defined in our axiomatization is the class of all chain coordinate systems of the form $O_i^w(a)$. In the previous section we spoke of the widest coordinate class which includes non-systematic coordinate systems; but in this section we ignore the widest class and apply CInv directly to the general class defined above. The first indication of discrimination within the general class appears in axiom 9 where we require the existence of at least one chain coordinate system $O_i^w(a)$ at every point **a** such that the metric function S_a^2 is differentiable in that system. The coordinate systems which axiom 9 requires to exist are called metrical. It is necessary to axioms 10 and 11 that at least one metrical system is available at each point. If it were possible to claim that every chain coordinate system is metrical, then there would be no axiomatic discrimination within the general coordinate class and CInv would be technically satisfied. Therefore, it is important to establish that the class of metrical systems is indeed a proper subclass of the general coordinate class.

Consider some chain $C_i(mn)$ which serves as the basis of a metrical chain coordinate system $O_i^w(a)$. Chain $C_i(mn)$ might be a geodesic path

going from point **m** to **n** through point **a** in figure 9. Now, consider another chain $C_j(mo)$ which coincides with $C_i(mn)$ from **m** up to and including point **a**, but which breaks sharply along some new geodesic path from **a** to point **o**. Surely we cannot expect the metric functions $S_a{}^2$ to be differentiable relative to a chain coordinate system $O_j{}^u(a)$ if it is differentiable relative to $O_i{}^w(a)$; and so axiom 9 does discriminate within the general class of chain coordinates defined in this axiomatization.

It is well known that the equations of general relativity are not generally covariant. They are restricted to coordinate systems which are connected by the group of real, continuous, and differentiable transformations such that the Jacobian is non-zero at all points.[33] These transformation restrictions have not concerned us directly, for we have found a way to deal with the requirement of CInv without being limited by the scope of Riemannian calculus. Instead, we have chosen to remain within this axiomatization, and our doing so has permitted us to ignore all of the above restrictions within the exception of the differentiability condition. For example, we did not need to discriminate against the class of discontinuous coordinates because such a class was not defined. One might hope to satisfy formally CInv by getting around the differentiability difficulty in the same way, that is, one might hope to avoid the initial introduction of a general class $O_i{}^w(a)$ by the construction given in Part II, and require only the existence of a metrical class with the desired differentiability properties. The metrical class would then be the general class, and so no axiomatic discrimination within the class would be necessary. One difficulty with this approach is that it requires an exhaustive specification of the properties of the metrical class independent of the features of the established metrical topological model constructed on the basis of axioms 1–8. It is desirable to define coordinates relative to such a model rather than stipulate their properties independently; for failure to define relative to the model would not only destroy the attempt to achieve a unification of the space properties, but it would also compound the offense to ~CP. The latter offense follows because the proposed program would imply the primitive status of the metrical coordinate class.

Before concluding this section, another possibility should be reviewed which might satisfy the formal requirement of invariance. It is natural to attempt a definition of "geodesic" in the metrical language and to construct only geodesic chain coordinate systems directly from those special chains. Axiom 9 might then state that geodesic chain coordinate systems are also metrical, thereby extending axiomatic reference to the general coordinate class as required by CInv. Although this method is a possible one it is not entirely straightforward, for our construction of a coordinate system involved a choice ω of coordinates θ and φ. This choice cannot be made uniquely on the basis of a single chain, so a choice would have to be made axiomatically to insure that the chosen parameters do indeed lead to a metrical coordinate system. This means either that the choice parameters must come from a primitive class, or that a wider class must first be defined. Neither of these alternatives is very satisfactory; for the first violates ~CP and the second violates CInv. An escape is to change the construction so as to avoid reference to parameters θ and φ and require metrical differentiability relative only to the class of coordinates r and t accompanying the class of all geodesic chains. A requirement of this kind should have the desired effect even though it would be very difficult to establish theorems on this basis. However these difficulties may be, to appeal to a special sub-class of chains in order to avoid axiomatic discrimination among coordinate systems is only to push the central problem back on the class of chains. In order to conform to CInv, we risk a violation of GInv as it applies to chains. It is certainly not obvious that we should care about discrimination among chains and one might claim that a proper subclass of chains is in fact preferred in nature. Perhaps non-geodesic chains should be discriminated against or not introduced at all. What is needed is a separate evaluation of the worth of the notion of geodesic paths in physics.

The non-linear Einstein field equations have the property that field defined masses will follow geodesic paths.[34] This circumstance has been of great interest because it has directed attention away from the notion of force which was so important in Newton's physics. The idea of a geodesic path is therefore rather central to general relativity. However, our evaluation of

[33] Bergmann, Peter G., *Introduction to the theory of relativity*, fifth ed., 60, New York, Prentice-Hall, 1950.

[34] Rainich, G. Y., *Mathematics of relativity*, 146, New York, Wiley, 1950.

it must proceed with more than relativity in mind, for we know that some of the ideas which seem natural to the theory will surely have to be abandoned. We will not attempt to evaluate the notion of "chain" as it may compare to "point" or "displacement," but ask instead if there is merit in continued discrimination within the class of chains when quantum physics is consulted. Probably the most direct answer to this question is given in Feynman's formulation of non-relativistic quantum mechanics.[35] To calculate the probability P_{ab} that a particle will appear at two points **a** and **b** in Minkowski Space (where $\mathbf{ab} < 0$), Feynman requires the probability that the particle will go along a certain path to be dependent on the classical action calculated for that path. The form of the dependence is an exponential whose imaginary phase is the classical action. The total contribution from *all* paths (i.e., all chains between **a** and **b**) is equated to the probability P_{ab}. By extending the method, Feynman shows that the total contribution reaching a given point in Minkowski Space is a function $\Psi(x, t)$ which satisfies Schrödinger's equation. The important point here is that all paths are given equal weight, and that there is no preference given to geodesics or even smooth differentiable curves. This formulation of quantum mechanics helps to focus the sense of our doubts about the extensibility of geodesics into the quantum domain. It seems risky to depend too heavily on geodesic chains when attempting to isolate vital notions of general relativity; so we determine that the use of special chains for the purpose of constructing a general coordinate class will constitute a violation of GInv applied to chains, and cannot be a very convincing way of avoiding a violation of GInv applied to coordinate systems.

We have now surveyed our violation of ∼CP and CInv in this axiomatization and presented some possible alternatives to our method. The consideration of alternatives was certainly not exhaustive, but it should help to clarify the character of the logical difficulty which we have attempted to overcome. It appears that ∼CP (denial of coordinate priority) will be violated so long as fundamental physical significance is given to the relativistic tensors g_{kl} and/or T_{kl}; for these quantities cannot be expressed in the widest class of coordinates. The core of the problem as it concerns both principles ∼CP and CInv is the need to guarantee metrical differentiability.

[35] Feynman, R. P., Space-time approach to non-relativistic quantum mechanics, *Rev. Mod. Phys.* **20**: 367–387, 1948.

IV. PROJECTION

There are two objectives which direct the following proposed extension of the theory. They are: (1) a satisfactory formal and conceptual unification of our notions of space, time and matter, and (2) the unqualified elimination of a restricted coordinate class. One can never fully justify objectives of this kind, but in support of them we will argue first that the recognition of a fundamental relation between space, time and matter seems to be demanded on the basis of classical relativity alone, and that the intelligibility and non-arbitrariness of this relation is the content of the unification idea whether expressed as a conclusion or a guide in research. Secondly, it is compelling to believe that coordinate systems function logically as name-plates which only identify events for the convenience of the physicist, and can therefore have nothing to do with the formulation of relationships holding between those events.

We will use our analysis of the relativity theory for the purpose of looking for a completely general approach to the above objectives. Since the unification view of both Clifford and Mach is compatible with these objectives our choice between the two is prompted largely by the ease with which it provides a direction to the program. In this final Part we therefore begin by outlining the characteristics of a theory which satisfies the principle of relativity expressed in ∼CP and CInv, and then discuss the appropriateness of our basic language to that kind of theory.

A physical theory which relies on differential expressions for the formulation of restrictive propositions is at once limited to a special coordinate class. This follows because the derivative of a function cannot be consistently defined relative to an arbitrary coordinate system. It cannot even be defined with the help of another function unless, as we have seen, an independent notion of geodesic is possible. For this reason, we are led rather directly to the conclusion that: *it must be required of a theory which fully affirms the principle of relativity that its fundamental propositions are not expressed by differential equations.*

Our criticism of the present axiomatization so far has been directed to the inadequacy of axiom

9. However, if we did have a theory such that the fundamental propositions could be stated independently of differential quantities, then axioms similar to 10 and 11 would have to be rejected along with axioms of metrical differentiability. This follows because axioms 10 and 11 employ the differential coefficients (tensor) g_{kl} and T_{kl}. Therefore, two questions arise immediately concerning the role of tensors g_{kl} and T_{kl} in a theory of the proposed kind. (1) How are we to impose sufficient axiomatic restriction on the chosen primitives without using arrays such as g_{kl} and T_{kl}, and (2) how is the proposed theory related to the tensor quantities (especially T_{kl}) which are in turn associated with laboratory operations under various agreed rules of correspondence?

(1) To answer the first question we will confine our attention to the use of tensor T_{kl} in axiom 11. To state the axiom we defined the array T_{kl} relative to the metrical coordinate class, and then effectively distinguished between real and fictitious matter. This distinction is again as follows: matter is real if the T_{44} component of the matter tensor is positive, and it is fictitious if T_{44} is either negative or zero while other components are non-zero. Axiom 11 then requires that there is no fictitious matter in the universe. This requirement has the effect of imposing strong restrictions on the metrical displacement assignments. The restrictions are precisely those imposed by the Einstein field equation as that equation is employed in general practice. Therefore, we are faced with the problem of deciding what to do with the apparently important content of axiom 11 in a theory which does not significantly rely on differential forms.

First, the distinction between real and fictitious matter may be questioned, and with it the advisability of axiom 11. If the field equation is not needed as an axiomatic restriction, it might then be viewed as a definition which simply relates the theory to classical quantities. Aside from noting this possibility, we cannot really decide very much without a better understanding of the needs of elementary particle theory. It seems advisable therefore to proceed on the assumption that a requirement comparable to axiom 11 must be retained. We suppose that the distinction between real and fictitious matter is a good one and that fictitious matter may be removed in a way that does not require the direct use of coordinates. This could mean the formulation of sufficiently strong propositions in the displacement language (i.e., the language used in this axiomatization) to impose constraints on the metric comparable to, although certainly not identical with, the requirement expressed in axiom 11.

Perhaps the above proposal will seem a bit more concrete if we recall that prior to the introduction of axiom 9, we were able to specify a great deal about the characteristics of our model of general Minkowski Space independent of a special coordinate class. It is tempting to consider the possibility of completing an axiomatic account of general relativity (or a modified general relativity) in a language similar to that used in axioms 1 through 8. In the present case we had to add axioms 9, 10, and 11 which are stated in a preferred coordinate language, and all three of these axioms impose strong restrictions on the primitive metrical displacements. However, it is reasonable to maintain that axioms 9 and 10 serve no purpose other than to prepare the way for axiom 11. Hence the importance of deciding the physical relevance of axiom 11. It is proposed above that one might possibly find a substitute language which can express that which remains essential in axiom 11, but which eliminates the need to select a preferred class of coordinates. It is not possible that the literal content of axiom 11 can be expressed independently of a select coordinate class, but then we do not suppose that the axiom is literally valid.

(2) In the second question above we asked how a theory of the proposed kind could be related to the tensor quantities (especially T_{kl}) which are in turn associated with laboratory operations under various agreed rules of correspondence. The question is epistemological, and our way of answering it will require some introductory remarks.

A physical theory or scientific formalism can be presented without giving any clear indication of its relevance to experience; and so we insist that rules of correspondence be given to complement the formalism and indicate which of its terms refer to something empirical. Margenau calls the formal terms "constructs" as indeed they are constructions from the formal apparatus of mathematics and logic.[36] We have used the word "model" in some instances to suggest a construct of particular interest. Constructs are denoted as circles in the diagram of figure 10,

[36] Margenau, H., *op. cit.*, chapter 5.

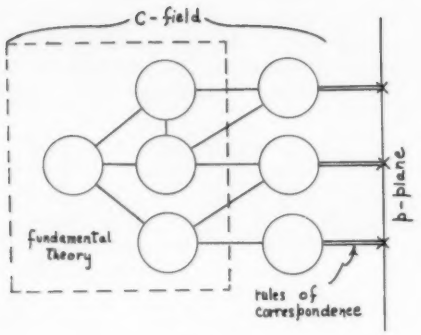


FIG. 10.

and all of the constructs which are relevant to some theory together comprise the *c*-field. Also in that diagram we have drawn a line on the right called the *p*-plane which represents the total of our empirical experiences. The double lines connecting some of the constructs with starred points on the *p*-plane are the rules of correspondence. They identify those terms or constructions which are representative of some aspect of *p*-plane experience. It is not in general possible to connect directly all of the constructs in the *c*-field with a single aspect of experience under a rule of correspondence. Those constructs which cannot be so connected are of interest in so far as they are needed to complete the formal postulational basis of scientific theory. The four constructs in figure 10 which are enclosed in a dotted line are here called the fundamental theory. The fundamental theory is the total of its member constructs and is a complete formal structure such as our axiomatization. Member constructs of the theory may be primitive terms or defined quantities, or perhaps complex models of special importance to the theory.

Since the constructs of the fundamental theory are not always connected directly to the *p*-plane under rules of correspondence, it is desirable to define other constructs which have more immediate empirical significance. They are shown in figure 10 as the three circles between the fundamental theory and the *p*-plane. These constructs are not necessary to the completeness of the fundamental theory, but they do follow deductively from it and are needed for the purpose of confirmation. Among them are those constructs which can be satisfactorily defined only when certain special conditions prevail. For instance, the classical trajectory of a particle is a construct which does not have very clear meaning in general but it can be reasonably defined when the particle considered is very heavy. That is, in the macroscopic limit as quantum mechanical effects may be ignored, a great many classical constructs are definable and have an established relation to the *p*-plane under some rule of correspondence. Constructs of this kind will be called limit constructs.

Now, hypothesize that the fundamental postulates of physics can be formulated without the introduction of a special coordinate class. Clearly we mean to suggest that these propositions extend beyond the theoretical domain of relativity and include restrictions strong enough to guarantee a satisfactory particle theory. Then if a special coordinate class is somehow specified, we could not expect that functions defined in the fundamental theory would be differentiable relative to that class; for a guarantee of this kind would require fundamental propositions which do discriminate within the widest coordinate class contrary to hypothesis. Conversely, since the solutions of a differential equation refer to a special coordinate class, it follows that a proposition of the form of a differential equation cannot correspond rigorously to a proposition in the hypothetical fundamental language. However, the correspondence may be statistical as in the case of a classical trajectory. For this reason, we refer to all special coordinate classes together with all terms which operate in differential propositions as limit constructs. The case is schematized in figure 11. We there characterize a special coordinate class and the matter tensor T_{kl} as limit constructs and indicate

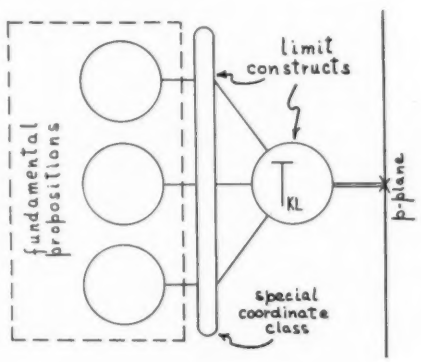


FIG. 11.

rules of correspondence from the tensor T_{kl} to the p-plane of experience.[37]

Our answer to the second question is, therefore, that the tensor quantities g_{kl} and T_{kl} are limit constructs which must relate statistically to the fundamental postulates of a theory of the proposed kind. It appears that the confirmation of such a postulational group would have to be indirect and possibly difficult. However, if the fundamental propositions are valid ones, then we can expect the limit propositions to coincide with the known probability formulation of microscopic physics. This means that we must anticipate a proof that statistically derived functions do exist in the limit language which correspond to experimentally successful quantum mechanical wave functions.

In this section we have so far followed the literal consequences of the principle of relativity and attempted to specify some of the characteristics which a theory must have to satisfy fully that principle. It has been found that functions defined in some fundamental language cannot be precisely translated into a differential propositional form, and that accuracy in the differential language must be regarded as somewhat fortuitous. We say that if the principle of relativity is rigorously valid in nature, then it is likely that physicists have had to accept probability as a primitive notion in quantum mechanics because the mechanics of the microcosm has been incorrectly formulated in differential form. As a means of giving direction to an investigation of relations which are completely relativistic, it is proposed above that a language similar to the one used in this axiomatization be considered as a vehicle of fundamental postulation.

The general form of, the requirements on, and the advantages of a theory of the proposed kind should be clear from what has preceded. It would take the form of a group of set theoretical propositions which deal directly with the constitutive elements of space-time. In a sense present theories already include a series of preliminary statements of this kind referring to the dimension and continuum properties of space, but they appear always as a group of assumptions which are incidental to the theory. The differential equation is generally assumed to be endowed with major physical content. In accordance with this proposal, however, these equations would be replaced by, and ultimately derived from propositions of a kind that permit the entire postulational group to be expressed in the same mathematical language.

[37] As noted previously, the T_{kl} tensor does not correspond directly to anything given in the p-plane, but the simplification here is not harmful or relevant to the proposal.

APPENDIX

THE AXIOMATIZATION

Primitives: 1. Point 2. Displacement Function

Points are represented by boldface letters **a**, **b**, **c**...**z**; and the displacement function assigns a real number (positive, negative, or zero) to every ordered pair of points in the universal set U. For two points **a** and **b**, the assignment to the ordered pair (ab) is called the displacement from point **a** to point **b**, and is represented by **ab**. The value of the assignment is assumed to be arbitrary apart from the requirements of the following axioms, and contingent limitations or boundary conditions we may wish to stipulate consistent with these axioms.

Those terms which have a special technical meaning in this axiomatization are introduced by definition and labeled **dfn**: Standard terminology is given in discussion, and those terms which have a technical character are underlined. Much of the terminology in the following discussion may be found in Kelley, John L., *General Topology*, New York, D. Van Nostrand Co.

Dfn: Special sets are (1) the universal set U of points, (2) the set N of real numbers, (3) the subset X of N which includes 0 and 1, plus all numbers between 0 and 1, and (4) the null set Λ.

A topological space is given by naming the set A of elements (i.e., points, numbers) which constitute the space, and the family $\mathbf{T}$ of *open sets* which constitute the topology. Every topology is subject to the requirement that (1) the intersection of members of $\mathbf{T}$ is a member of $\mathbf{T}$, and (2) the union of members of $\mathbf{T}$ is a member of $\mathbf{T}$. If the set A is open (i.e., if A is a member of $\mathbf{T}$), then the space is designated by the pair $(A, \mathbf{T})$. The complement $A-V$ of an open set $V \epsilon \mathbf{T}$ is called a *closed set*.

Given a space $(A, \mathbf{T})$, and a subset $Q \epsilon A$, the subspace $(Q, \mathbf{T})$ is a topological space with a topology (called the *relativized topology*) formed by the intersection of Q with open sets in $\mathbf{T}$. For instance, if we designate the *usual topology*

of the real axis by $\mathbf{T}_0$, then the pair $(N, \mathbf{T}_0)$ specifies the topological space which we understand to be an infinitely extended one dimensional continuum. An open set in this case is a set which may be formed by the union of any collection of open intervals. The subspace $(X, \mathbf{T}_0)$ employs the same topology confined to the interval 0 to 1. Set X is both open and closed in the space $(X, \mathbf{T}_0)$. In general, set A in the space $(A, \mathbf{T})$ is both open and closed; and property which it shares with the null set, but no other subset of itself.

A family $\widehat{B}$[38] of sets is a *base for* $\mathbf{T}$ iff each member of $\mathbf{T}$ is a union of members of $\widehat{B}$. The usual base of the real axis is the family of all open intervals on the real axis.

A family $\widehat{SB}$ of sets is a *subbase for* $\mathbf{T}$ iff the family of finite intersections of members of $\widehat{SB}$ is a base for $\mathbf{T}$. That is, $\widehat{SB}$ is a subbase for $\mathbf{T}$ iff each member of $\mathbf{T}$ is the union of finite intersections of members of $\widehat{SB}$.

The *neighborhood of an element* $p \epsilon A$ relative to a space $(A, \mathbf{T})$ is a set P such that there exists an open subset $V \subseteq P$ to which p belongs. That is, if $p \epsilon V$, $V \epsilon \mathbf{T}$, and $V \subseteq P$, then P is a neighborhood of p. Conversely, an element p of a subset P of a topological space is an *interior element* of P iff subset P is a neighborhood of p. We will say that a space $(A, \mathbf{T})$ is *Hausdorff* iff, whenever p and q are distinct elements of the space, there exists disjoint neighborhoods of p and q. It follows from these definitions that, if a space is Hausdorff, then for any element p of the space, the set $\{p\}$ is closed. To show this, consider some point q in the complement $A - \{p\}$ of set $\{p\}$. The Hausdorff condition requires that there exists a neighborhood, and hence an open set V containing q which does not contain point p. The set V is therefore confined to set $A - \{p\}$ so that we may write $A - \{p\}$ as the union of such open sets. The set $A - \{p\}$ is therefore open, and so $\{p\}$ is closed.

A *limit element* of the subset P of a topological space is an element p such that every neighborhood of p contains an element of the set $P - \{p\}$. Designate: $p \lim P$.

The *boundary* of a subset P of a topological space $(A, \mathbf{T})$ is the set of all elements of A which are interior to neither P nor the complement of P (i.e., A–P). Evidently, element p is in the boundary of P iff it is a limit element of both P and A–P. Designate: $p \operatorname{bdr} P$. From these definitions it follows that the boundary of P will be disjoint from P if P is an open set; and it will be contained in P if P is a closed set.

A *function* f is a relation between the elements y of a set A, and the elements θ of a set B. The function assigns values, or attaches the elements of B to elements of A, where set A is called the *domain*, and set B the *range* of f. Designate: $\theta = f(y)$. A function specifies those members of B which are an *image*, a *mab* or *transform* of members of A.[39] The image under f of the subset $Y \subset A$ is a subset $L \subset B$, represented by $L = f[Y]$. The inverse f^{-1} of the function f is a function which transforms elements p (of the range of f) into the elements θ (of the domain of f), and is represented by $\theta = f^{-1}(p)$.

A function $f(\theta) = p$ is *single valued* if it chooses one and only one range element p for every element θ of its domain. If the inverse of f is single valued, then for every θ in the domain of f: $f^{-1}[f(\theta)] = \theta$. In the case a function is single valued, and its inverse is also single valued, the mapping is said to be *one-to-one*. A one-to-one function associates a unique member of its range to each member of the domain, and a unique member of its domain to each member of the range.

Consider two topological spaces $(A, \mathbf{T}_a)$ and $(B, \mathbf{T}_b)$, where elements of the set A are denoted by lower case Greek letters and elements of the set B are denoted by lower case Roman letters. If a function f is specified with a domain A and a range B, then we may inquire whether the subsets of A which transform under the inverse mapping from open sets in $(B, \mathbf{T}_b)$, are open sets in $(A, \mathbf{T}_a)$. If they are, then we say that the map f is continuous. A continuous function is characterized in figure 12 where topological space $(A, \mathbf{T}_a)$ is mapped into $(B, \mathbf{T}_b)$ by a function f. The subset $\{\beta, \gamma\}$ of set A is an open set in $\mathbf{T}_a$ (i.e., it is a member of the family $\mathbf{T}_a$), and the subset $\{n, r\}$ of B is an open set in $\mathbf{T}_b$. The function f is continuous because every open set in $\mathbf{T}_b$ such as $\{n, r\}$ transforms under the inverse mapping to an open set in $\mathbf{T}_a$. In figure 12, we let any uninterrupted sequence of elements along either line A or B be an open set.

[38] Families of sets other than $\mathbf{T}_1$ will be designated with a capital letter plus the symbol (^).

[39] It has been required that the primitive displacement function identify a number n for every ordered pair of points. This function could have been represented by $n = d(a, b)$.

For example, $\{\alpha, \beta\}$ and $\{\beta, \gamma, \delta\}$ are open in $\mathbf{T}_a$. Also, $\{m, n, r, s\}$ and $\{r, s, t\}$ are open in $\mathbf{T}_b$.

An example of a continuous function is one which maps a subset of points of a two-dimensional plane into every point of a three dimensional cube such that the inverse mapping is a projection of the cube into the plane. Let the space of the plane be represented by $(A, \mathbf{T}_a)$, and the space of the cube by $(B, \mathbf{T}_b)$ in figure 12. In this case, we require that the inverse of the function (i.e., the projection) be single valued. If β is the projection of r into the plane, then we expect that the image of any neighborhood of r be a neighborhood of β. In the sense of the usual topology of a plane and a cube, this is accomplished by requiring that every open set in the cube topology $\mathbf{T}_b$ map into an open set in the plane topology $\mathbf{T}_a$. The function f is therefore continuous although it is not single valued.

Examples of functions which are both continuous and single valued are well known in algebra. If given a continuous single valued function $y = f(x)$, where y represents a real numerical value assigned to points x of some domain set, then the more familiar meaning of "continuous function" follows from the above definition; that is, for every point x_0, and for any real positive number ϵ, there exists a neighborhood of x_0 such that the value of the function at every point in the neighborhood is within a radius ϵ of the value $y_0 = f(x_0)$. To prove that such a neighborhood exists for every radius ϵ, select any open interval on the y-axis which includes y_0, but excludes values outside of the desired radius. This open interval is an open set in the usual topology of the real y-axis, and is required to transform into an open set of the domain under the inverse mapping f^{-1}. The selected image set is therefore a neighborhood of x_0 which, because f is single valued, excludes points having values assigned outside of the chosen ϵ radius of y_0.

Two topological spaces are *homeomorphic* iff there exists a one-to-one function, called a homeomorphism, which transforms one of the spaces into the other, and such that the function and its inverse are continuous.

It is shown in Sierpinski's *General Topology*, p. 22, that limit elements transform into corresponding limit elements under a homeomorphism f. We therefore state a known theorem. In the following, theorems will be designated **thm**.

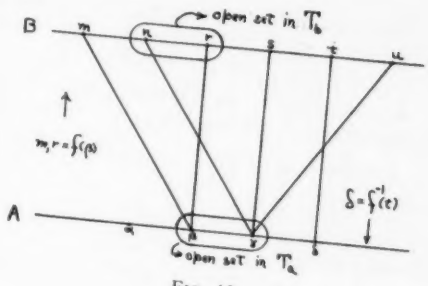


FIG. 12.

Thm 0: If there exists a homeomorphism f from a space $(A, \mathbf{T}_a)$ to a space $(B, \mathbf{T}_b)$, and if an element p of A is a limit element of any subset $P \subseteq A$, then the image $f(p)$ of p is a limit element of the image subset $f[P] \subseteq B$.

Turning attention now to the universal set U of primitive points, we construct a topology for U by first defining a set called the basic set.

Dfn: For any two points **a** and **b**, the *basic set* specified by **a** and **b** is the set of all points **c** such that $\mathbf{ac} < 0$ and $\mathbf{cb} < 0$. Designate: $B(ab)$.

Dfn: The *subbase family of the point set* is the family of all basic sets B. From our previous definition, it follows that the family of all finite intersections of basic sets is the *base family of the point set*.

Dfn: The union of any collection of members of the base family of the point set is an *open set* of the point topology. The family of all such open sets is the *topology* $\mathbf{T}_1$ *of the point set*.

Dfn: A *neighborhood of a point* is any point set Q such that there exists an open subset $V \subseteq Q$ to which the point belongs. That is, if $\mathbf{c} \epsilon V$, $V \epsilon \mathbf{T}_1$, and $V \subseteq Q$, then set Q is a neighborhood of **c**. Designate: $N(c)$.

Points will not be boldface when occurring in bracketed expressions such as $B(ab)$ and $N(c)$. The set $B(ab)$ is uniquely specified by points **a** and **b**, but particular sets may also be specified by subscripts such as B_1 or $N_1(c)$. It follows from these definitions that every basic set is a neighborhood of every one of its members (i.e., every member of B_1 is interior to B_1); but that a neighborhood is not necessarily a basic set. Therefore, although we may sometimes refer to a basic set as a neighborhood, a neighborhood will not be identified directly with some basic set $B(xy)$.

Dfn: The *successor set* of a point a is the set of all points **b** such that $\mathbf{ab} < 0$. Designate: $S(a)$.

Dfn: A displacement **ab** is called *successor directed* from **a** to **b** iff **b** lim $S(a)$. Designate: $\overrightarrow{\mathbf{ab}}$.

Dfn: A displacement **ab** is called *directed* iff $\overrightarrow{\mathbf{ab}}$ or $\overrightarrow{\mathbf{ba}}$.

Dfn: For a given point **a**, the *relative distance function* $d_a^{(1)}(u)$ is defined on the domain U, and assigns a number to every point **u** such that $d_a^{(1)}(u) = \mathbf{ua}$. The *relative distance function* $d_a^{(2)}(u)$ assigns a number to every point **u** such that $d_a^{(2)}(u) = \mathbf{ua}$.

Dfn: If Q is a subset of U, and the space (Q, T_1) is homeomorphic with the space (X, T_0), then there exists a homeomorphism **f** which maps 0 into a point $\mathbf{a} \epsilon Q$, and 1 into a point $\mathbf{b} \epsilon Q$. Accordingly, points **a** and **b** are called end points of Q relative to f. The remaining points are images of the interior members of X, and are called *included points* of Q relative to f.

Dfn: A displacement **ab** is timelike either if $\mathbf{ab} < 0$, or $\mathbf{ba} < 0$. Displacements which are not timelike are called *spacelike*.

Dfn: A set of points Q is a chain if (1) the displacement between any two members of Q is timelike, and if (2) there exists a homeomorphism f from the space (Q, T_1) to the space (X, T_0). Designate: $C^f(ab)$, where **a** and **b** are end points of the homeomorphism f.

The axioms will be presented in several groups. Group I contains axioms which provide for an ordering relation among non-positive displacements, continuity, and the existence of chains:

GROUP I:

Axiom 1: The space (U, T_1) is Hausdorff.

Axiom 2: For all points **a** and **b**: $|\mathbf{ab}| = |\mathbf{ba}|$.

Axiom 3: For all points **a**, **b**, and **c**; if $\overrightarrow{\mathbf{ab}}$ and $\overrightarrow{\mathbf{bc}}$, then $\mathbf{ac} \leqslant \mathbf{ab} + \mathbf{bc}$.

Axiom 4: For all points **a** and **b** such that $\mathbf{ab} = 0$, the displacement **ab** is directed.

Axiom 5: For every point **a**, and all points $u \epsilon U$, the relative distance functions $d_a^{(1)}(u)$ and $d_a^{(2)}(u)$ are continuous and single valued.

Axiom 6: For all points **a** and **b** such that $\mathbf{ab} < 0$, there exists at least one chain $C^f(ab)$.

Thm 1: If $\mathbf{ab} < 0$, then $\overrightarrow{\mathbf{ab}}$.

Proof: Hypothesize that $\mathbf{ab} < 0$ and that there exists a neighborhood $N_0(b)$ of **b** such that for all $\mathbf{c} \epsilon N_0(b) - \{\mathbf{b}\}$, $\mathbf{ac} \geqslant 0$. If we represent the negative numbers by N^-, then N^- is an open set in (N, T_0); and so the set $V = d_a^{(1)-1}[N^-]$ is open in (U, T_1) in accordance with axiom 5. By hypothesis, no member of $N_0 - \{\mathbf{b}\}$ is contained in V; therefore, $N_0 \bigcap V = \{\mathbf{b}\}$. Since N_0 and V are open sets in T_1, the set $\{\mathbf{b}\}$ must be open contrary to axiom 1. The hypothesis is contradicted, and it follows that all $N(b)$ contain points **c** such that $\mathbf{ac} < 0$. Therefore, **b** lim $S(a)$ and by definition $\overrightarrow{\mathbf{ab}}$.

Corollary 1: If $\mathbf{ab} < 0$ and $\mathbf{bc} < 0$, then $\mathbf{ac} \leqslant \mathbf{ab} + \mathbf{bc}$.

Proof: From thm 1, $\overrightarrow{\mathbf{ab}}$ and $\overrightarrow{\mathbf{bc}}$. From axiom 3, it follows directly that $\mathbf{ac} \leqslant \mathbf{ab} + \mathbf{bc}$.

Thm 2: If $\mathbf{ab} < 0$, then $\mathbf{ba} = -\mathbf{ab}$.

Proof: Axiom 2 requires that $|\mathbf{ba}| = |\mathbf{ab}|$. Suppose $\mathbf{ba} = \mathbf{ab} < 0$, then according to corollary 1, $\mathbf{aa'} \leqslant \mathbf{ab} + \mathbf{ba'}$, where **a** and **a'** are the same point. But then the displacement from point **a** to itself is negative. In that case, $\mathbf{aa''} \leqslant \mathbf{aa'} + \mathbf{a'a''}$ (where **a**, **a'**, and **a''** are again the same point) in accordance with corollary 1, and contrary to axiom 5 which requires the single valuedness of the function $d_a^{(1)}(a)$. The supposition $\mathbf{ba} = \mathbf{ab} < 0$ is therefore false. Hence, $\mathbf{ba} = -\mathbf{ab}$.

Dfn: If $\mathbf{ab} < 0$, then point **a** is called a *predecessor* of point **b**, and point **b** is called a *successor* of point **a**.

Dfn: The *predecessor set* of a point **a** is the set of all points **b** such that $\mathbf{ba} < 0$. Designate: $P(a)$.

From theorem 2, it follows that all timelike displacements are asymmetric in the relation "successor of" or "predecessor of," such that the displacement from any point to a successor is negative, and the reverse displacement is positive. In addition, corollary 1 requires that the relation "successor of" or "predecessor of" is transitive. It follows that a chain is a set of points which is ordered by a connected, transitive, and asymmetric relation recalling a unidirectional string of points. The homeomorphism requirement in the definition of chain is

intended to say that this string is also a continuum.

We now require that the designation of a chain $C^f(ab)$ reflect the asymmetry of end points **a** and **b**. When **a** appears first in the brackets, it will mean that **a** is a predecessor of **b**. In addition, we may occasionally designate a chain by $C^f(abcd)$, where $\mathbf{ab} < 0$, $\mathbf{bc} < 0$, and $\mathbf{cd} < 0$, such that **b** and **c** are included points of a chain with end points **a** and **b** relative to f.

Theorem 2 permits the further observation that for all points **a** and **b**, neither **a** nor **b** are members of $B(ab)$; for no point can be successor or predecessor of itself.

Thm 3a: If **p** is a limit point, and is disjoint from the successor set of point **a**, then **ap** is spacelike.

Proof: If $\mathbf{p}$ not ϵ $S(a)$; $\mathbf{p}$ lim $S(a)$, then **ap** is not negative, and all $N(p)$ contain points **m** such that $\mathbf{am} < 0$. Since the function $d_a^{(1)}(u)$ is continuous at point **b**, it follows that $\mathbf{ap} = 0$. But then $\mathbf{pa} = 0$ by axiom 2. The displacement **ap** is therefore not timelike, hence spacelike.

Thm 3b: If **q** is a limit point, and is disjoint from the predecessor set of point **b**, then **bq** is spacelike.

Proof: If $\mathbf{q}$ not ϵ $P(b)$; $\mathbf{q}$ lim $P(b)$, then **qb** is not negative, and all $N(q)$ contain points n such that $\mathbf{nb} < 0$. Hence by theorem 2, all $N(q)$ contain points such that $\mathbf{bn} > 0$. Since the function $d_b^{(1)}(u)$ is continuous at point **q**, it follows that **bq** is not negative. The displacement **bq** is therefore not timelike, hence spacelike.

Thm 3c: If **q** bdr $B(xy)$, then either **qx** is spacelike, or **qy** is spacelike.

Proof: If **q** is a boundary point of the open set $B(xy)$ then $\mathbf{q}$ lim $B(xy)$ and $\mathbf{q}$ not ϵ $B(xy)$. Since $B(xy) = S(x) \bigcap P(y)$, it follows that: (1) $\mathbf{q}$ lim $S(x)$, and (2) $\mathbf{q}$ lim $P(y)$, and (3) either $\mathbf{q}$ not ϵ $S(x)$ or $\mathbf{q}$ not ϵ $P(y)$. Therefore, theorems 3 *a*, *b* require that either **qx** is spacelike or **qy** is spacelike.

Lemma 1: If $\mathbf{x}, \mathbf{y} \epsilon C^f(ab)$, then $f[C^f(ab) \bigcap B(xy)]$ has two boundary points at the most, which are $f(x)$ and $f(y)$.

Proof: Suppose a number $f(q) \epsilon X$ other than $f(x)$ or $f(y)$ is a boundary element of the set $f[C^f(ab) \bigcap B(xy)]$. See figure 13. Then by

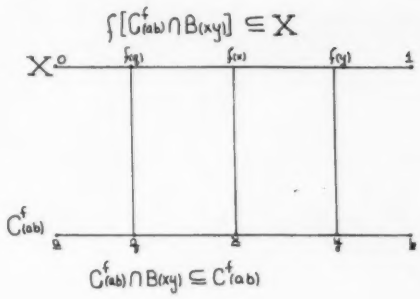


FIG. 13.

theorem 0 and the definition of boundary, point **q** is a boundary point of $C^f(ab) \bigcap B(xy)$. To be a boundary point of this open set,[40] **q** cannot be in $B(xy)$, for it is already in the set $C^f(ab)$. It follows that **q** bdr $B(xy)$. But then from theorem 3*c*, either **xq** is spacelike or **yq** is spacelike contrary to the requirement that these displacements are timelike (i.e., $\mathbf{x}, \mathbf{y}, \mathbf{q} \epsilon C^f(ab)$). The supposition that $f(q)$ is other than $f(x)$ and $f(y)$ is therefore false. It follows that the open set $f[C^f(ab) \bigcap B(xy)]$ can have two boundary points at the most, which are $f(x)$ and $f(y)$.

Lemma 2: For all $\mathbf{x}, \mathbf{y} \epsilon C^f(ab)$ such that $\mathbf{xy} < 0$, the set $C^f(ab) \bigcap B(xy)$ is non-null.

Proof: Hypothesize that: $\mathbf{x}, \mathbf{y} \epsilon C^f(ab)$, $\mathbf{xy} < 0$, and $C^f(ab) \bigcap B(xy) = \Lambda$. It follows that members of $C^f(ab) - \{\mathbf{x}, \mathbf{y}\}$ must be either in $P(x)$ or $S(y)$. The sets $M = P(x) + \{x\}$ and $N = S(y) + \{y\}$ contain all of the points in $C^f(ab)$, and so images $f[M]$ and $f[N]$ are complementary in X. Since neither M nor N is null, it follows from the known compactness of the set X, that there is a limit element of one of the images which is not a member of that image. Suppose it is $f[M]$ which has a limit element $f(q)$, where $\mathbf{q} \epsilon C^f(ab)$, such that $f(q)$ not ϵ $f[M]$. By theorem 0, **q** lim M. But then **q** lim $P(x)$, for there exists at least one $N(q)$ which excludes point **x** (i.e., the space is Hausdorff). Since furthermore **q** not ϵ $P(x)$, it follows from theorem 3*b* that **qx** is spacelike contrary to the requirement that **qx** is timelike (i.e., $\mathbf{x}, \mathbf{q} \epsilon C^f(ab)$). The supposition that $f(q)$ not ϵ $f[M]$, and $f(q)$ lim $f[M]$ is therefore false. Similarly, $f(q)$ not ϵ $f[N]$, and $f(q)$ lim $f[N]$ is

[40] $C^f(ab) \bigcap B(xy)$ is an open set in the space $(C^f(ab), \mathbf{T}_1)$ because it is the intersection of $C^f(ab)$ with a member of $\mathbf{T}_1$.

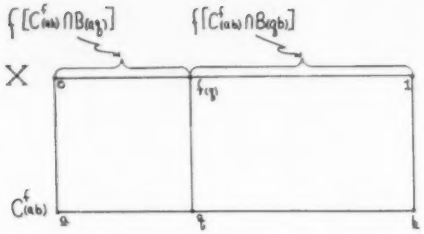


FIG. 14.

false. Hence, the hypothesis is contradicted, and the lemma is proved.

Thm 4: The set $C^f(ab)\bigcap B(ab)$ contains every included point of $C^f(ab)$.

Proof: From lemma 1, the open set

$$f[C^f(ab)\bigcap B(ab)]$$

can have two boundary points at the most, which are 0 and 1. Since $C^f(ab)\bigcap B(ab)$ does not contain **a** and **b**, it follows that the open set $f[C^f(ab)\bigcap B(ab)]$ does not contain 0 and 1. There are only two open sets in X which fulfill these conditions: (1) the null set, and (2) the open interval between 0 and 1. But $f[C^f(ab)\bigcap B(ab)]$ can be null only if $C^f(ab)\bigcap B(ab)$ is null, and lemma 2 requires that this set is non-null. Therefore, $f[C^f(ab)\bigcap B(ab)]$ must be identical with the open interval from 0 to 1. Hence, the inverse of this open set; $C^f(ab)\bigcap B(ab)$, contains every included point of $C^f(ab)$.

The above theorem establishes that the end points of the chain are also terminal points of the predecessor-successor relationship. They are the first and last points of the ordered chain set.

Lemma 3: If point **q** is included in the chain $C^f(ab)$, then the set $f[C^f(ab)\bigcap B(aq)]$ is identical with the open interval from 0 to $f(q)$, and $f[C^f(ab)\bigcap B(qb)]$ is identical with the open interval from $f(q)$ to 1.

Proof: From lemma 1. $f[C^f(ab)\bigcap B(aq)]$ has two boundary points at the most, which are 0 and $f(q)$. From lemma 2, $f[C^f(ab)\bigcap B(aq)]$ is non-null. Furthermore, the numbers 0 and 1 are not contained in $f[C^f(ab)\bigcap B(aq)]$ since **a** is not in $B(aq)$ and since by theorem 4 the point **b** cannot be in $B(aq)$. The only open set of X which can fulfill these requirements is the open interval between 0 and $f(q)$. See figure 14.

Similarly, the set $f[C^f(ab)\bigcap B(qb)]$ is non-null. It does not contain the numbers 0 and 1, and has two boundary points at the most which are $f(q)$ and 1. It is therefore the open interval between $f(q)$ and 1.

Lemma 3 makes the essential properties of the chain transparent. If $\mathbf{q}\epsilon C^f(ab)$, then all of the successors of **q** in the chain are mapped under f into numbers greater than $f(q)$ or less than $f(q)$, depending on whether f maps point **a** into 0 or 1. The next three theorems follow from lemma 3, and express the properties of the chain which will be most frequently used. In these, and in subsequent theorems, the particular homeomorphism f is not essential so long as we specify that a given chain $C(ab)$ is restricted to a unique set of included points. This will be accomplished by the use of numerical subscripts. That is, chains $C(ab)$, $C_1(ab)$, and $C_2(ab)$ represent different paths from point **a** to point **b**. It will, therefore, be convenient to drop superscript f in the designation of a chain. Furthermore, since the property of being an included member of a chain is not a function of the homeomorphism f, we use the symbol ϵ' to mean included membership. That is, $\mathbf{q}\epsilon' C(ab)$ means that **q** is an included point of $C(ab)$. On the other hand $\mathbf{q}\epsilon C(ab)$ will continue to mean membership in the usual sense.

Thm 5: If $\mathbf{x}, \mathbf{y}\epsilon C(ab)$, and $\mathbf{xy} < 0$, then the set $C(ab)\bigcap B(xy) + \{x, y\}$ is a chain.

Proof: The points in $C(ab)\bigcap B(xy)$ are those members of the chain which are successors of **x** and predecessors of **y**. If the homeomorphism is f, then by lemma 3, $f[C(ab)\bigcap B(xy)]$ is the open interval between $f(x)$ and $f(y)$. The closed interval between $f(x)$ and $f(y)$ is known to be homeomorphic with X. The homeomorphic relation is transitive; and so it follows that the point set $C(ab)\bigcap B(xy) + \{xy\}$ is homeomorphic with X. Furthermore, every displacement between members of $C(ab)\bigcap B(xy) + \{xy\}$ is timelike. Therefore, $C(ab)\bigcap B(xy) + \{xy\}$ is a chain.

Thm 6: The union $C(ab)\bigcap C(bc)$ is a chain $C(abc)$.

Proof: The set X is homeomorphic with the closed interval between 0 and $\frac{1}{2}$. There is,

therefore, a homeomorphism f, which maps $C(ab)$ into this interval such that $f_1(a) = 0$ and $f_1(b) = \frac{1}{2}$. Similarly, there exists a homeomorphism f_2 which maps $C(bc)$ into the closed interval from $\frac{1}{2}$ to 1 such that $f_2(b) = \frac{1}{2}$ and $f_2(c) = 1$. The function $f = f_1 + f_2$ therefore maps $C(ab) \bigcup C(bc)$ one-to-one into X. To show that f is a homeomorphism, we have only to insure that an open interval in X which does not contain 0 and 1 but which does contain $\frac{1}{2}$ is an open set in $C(ab) \bigcup C(bc)$, and that for points $\mathbf{x} \epsilon C(ab)$ and $\mathbf{y} \epsilon C(bc)$, the open set $[C(ab) \bigcup C(bc)] \bigcap B(xy)$[41] maps into an open set in X.

A. Let $f(x)$ and $f(y)$ be boundary elements of an open set in X such that $0 < f(x) < \frac{1}{2}$, $\frac{1}{2} < f(y) < 1$, and every number between $f(x)$ and $f(y)$ is in the set. From lemma 3, numbers between $f(x)$ and $\frac{1}{2}$ transform into successors of point $\mathbf{x}$ in $C(ab)$. The closed interval from $\frac{1}{2}$ to 1 transforms into point $\mathbf{b}$ plus all successors of $\mathbf{b}$ in $C(bc)$. These points are all successors of $\mathbf{x}$ by theorem 4 and corollary 1. The image of the open interval from $f(x)$ to and including 1, therefore, maps into successors of point $\mathbf{x}$ in the set $C(ab) \bigcup C(bc)$. Similarly, the half open interval from $f(y)$ to and including 0 maps into all of the predecessors of point $\mathbf{y}$ which are in $C(ab) \bigcap C(bc)$. The intersection of the two half open intervals in X is the open interval between $f(x)$ and $f(y)$, and maps into those members of $C(ab) \bigcup C(bc)$ which are both successors of $\mathbf{x}$ and predecessors of $\mathbf{y}$. It therefore identifies the set $[C(ab) \bigcup C(bc)] \bigcap B(xy)$, which is open.

B. Consider the open set

$$[C(ab) \bigcup C(bc)] \bigcap B(xy)$$

where $\mathbf{x} \epsilon' C(ab)$ and $\mathbf{y} \epsilon' C(bc)$. Under the homeomorphism f, all successors of $\mathbf{x}$ in $C(ab) - \{b\}$ transform into numbers between $f(x)$ and $\frac{1}{2}$ in accordance with lemma 3. Furthermore, members of $C(bc)$, which are all successors of $\mathbf{x}$ by theorem 4 and corollary 1, transform into the closed interval from $\frac{1}{2}$ to 1. All successors of $\mathbf{x}$ in $C(ab) \bigcup C(bc)$ therefore map into the half open interval from $f(x)$ to and including 1. Similarly, all predecessors of $\mathbf{y}$ map into the half open interval from $f(y)$ to and including 0. Since the set $[C(ab) \bigcup C(bc)] \bigcap B(xy)$ contains all points in $C(ab) \bigcup C(bc)$ which are both successors of $\mathbf{x}$ and predecessors of $\mathbf{y}$, its image in X will be the open set between $f(x)$ and $f(y)$.

[41] This is an open set in the space $(Q, \mathbf{T}_1)$ where $Q = C(ab) \bigcap C(bc)$, because it is the intersection of Q with a member of $\mathbf{T}_1$.

The subscript convention adopted above for chains will now have the more explicit meaning that for any point $\mathbf{c} \epsilon' C_1(ab)$:

$$C_1(ab) = C_1(acb) = C_1(ac) + C_1(cb).$$

Dfn: A *world chain* is the union of an infinite number of chains $\cdots + C_i(ab) + C_i(bc) + C_i(cd) + C_i(de) + \cdots$ such that (1) there is no member of the set which is a predecessor of every other member, and (2) there is no member of the set which is a successor of every other member, and (3) there is no limit point of the set which is not a member of the set. Designate: C_i.

Thm 7: If $\mathbf{q} \epsilon' C(ab)$, then every $N(q)$ contains a predecessor of $\mathbf{q}$, and a successor of $\mathbf{q}$ included in the chain.

Proof: Consider any $N(q)$. The image of $C(ab) \bigcap N(q)$ under the homeomorphism f will be the neighborhood $f[C(ab) \bigcap N(q)]$ in X containing $f(q)$ as an interior element. Therefore, the number $f(q)$ will not be a limit point of the complement of $f[C(ab) \bigcap N(q)]$. It follows that every neighborhood of $f(q)$ contains interior numbers less than $f(q)$ and greater than $f(q)$. In particular $f[C(ab) \bigcap N(q)]$ contains interior numbers less than $f(q)$ and greater than $f(q)$. By lemma 3, the set $C(ab) \bigcap N(q)$ must therefore contain successors of point $\mathbf{q}$ and predecessors of point $\mathbf{q}$.

Thm 8: If $\mathbf{c} \epsilon N(c) \bigcap B(ab)$, then there exists a predecessor $\mathbf{x}$ of $\mathbf{c}$ and a successor $\mathbf{y}$ of $\mathbf{c}$ such that $\mathbf{x}, \mathbf{y} \epsilon N(c) \bigcap B(ab)$.

Proof: If $\mathbf{ac} < 0$ and $\mathbf{cb} < 0$, then by axiom 6 and theorem 6, there exists a chain $C(abc)$. From theorem 7, there is an included point $\mathbf{p}$ of the chain which is a predecessor of $\mathbf{c}$ in the neighborhood $N(c)$. Theorem 4 requires that $\mathbf{ab} < 0$ and $\mathbf{pb} < 0$, so that $\mathbf{p} \epsilon B(ab)$. Therefore $\mathbf{p} \epsilon N(c) \bigcap B(ab)$. Furthermore, from theorem 7, there is an included point $\mathbf{q}$ of the chain which is a successor of $\mathbf{c}$ in the neighborhood $N(c)$. Theorem 4 requires that $\mathbf{aq} < 0$ and $\mathbf{qb} < 0$, so that $\mathbf{q} \epsilon B(ab)$. Therefore $\mathbf{q} \epsilon N(c) \bigcap B(ab)$.

Thm 9: For any neighborhood $N(c)$ of a point **c**, there is a predecessor and a successor of **c** contained in that neighborhood.

Proof: If $\mathbf{c}\epsilon N(c)$, then there is an open set V such that $\mathbf{c}\epsilon V$ and $V \subseteq N$. Since V is a union of intersections of basic sets, there exists a set $B(ab)$ such that point **c** is a member. From theorem 8, there exists a predecessor and successor of **c** in the intersection of $B(ab)$ with any other basic set $B(ef)$. Furthermore, there exists a predecessor and successor of **c** in the intersection of the neighborhood $B(ab) \bigcap B(ef)$ with any third basic set. Hence, any finite intersection of basic sets including **c** also includes a predecessor and successor of **c**. It follows that the open set V, and therefore the neighborhood $N(c)$ contain a predecessor and successor of point **c**.

Thm 10: For every point $\mathbf{c}\epsilon U$, there exists a predecessor of **c** and a successor of **c**.

Proof: Consider the function $d_c(c')$, where **c** and **c'** are the same point. According to axiom 5, this function is single valued and continuous, and assigns to point **c'** the number **cc'** from the real axis. There exists an open set V on the real axis containing the number **cc'**: and from the definition of continuous function, the inverse map $d_c^{(1)}[V^{-1}]$ of V is an open set containing point c. This open set is a neighborhood of **c**, and from theorem 9 contains a predecessor and a successor of point **c**.

The preceding theorem establishes the very desirable property that every point in the universe is included in a chain; or, we may say, it is temporally connected to "past" points and to "future" points. There can be, therefore, no first point or last point in the temporal sense. It is also clear from the previous theorems that for every neighborhood $N_1(c)$, there is a larger neighborhood $N_2(c)$ and a smaller neighborhood $N_0(c)$, that is, there exists a $N_0(c)$ and a $N_2(c)$ such that $N_0(c) \subset N_1(c) \subset N_2(c)$. Hence, point **c** has no largest or smallest neighborhood.

Thm 11: The displacement from any point to itself is zero.

Proof: If **a** and **a'** are the same point, we know from theorem 2 that **aa'** is not timelike because **aa'** = **a'a**. Suppose then that **aa'** = **a'a** is positive. In that case, the number $n = d_a^{(1)}(a')$ is positive, and there must exist at least one neighborhood $N_1(a')$ which excludes all points c such that **ac** < 0. But from theorem 9, there exists a point $\mathbf{q}\epsilon N_1(a')$ such that **a'q** < 0; and so **aq** < 0. The supposition that **aa'** = **a'a** > 0 is therefore false, and so the displacement **aa'** $= 0$.

In the following theorems it will be important to note that the successor set $S(a)$ of any point **a**, and the predecessor set $P(b)$ of any point **b**, are open sets. An open set is one which is equal to the union of a family of base sets, and any subfamily of the basic sets is such a family. Therefore, to prove the openness of $S(a)$ and $P(b)$, we need only show that:

$$S(a) = \bigcup[B(ap) : \mathbf{p}\epsilon S(a)];$$

and

$$P(b) = \bigcup[B(qb) : \mathbf{q}\epsilon P(b)].^{42}$$

That is, for any point $\mathbf{x}\epsilon S(a)$, theorem 10 requires the existence of a point **p** such that **xp** < 0. Therefore $\mathbf{x}\epsilon B(ap)$, and according to corollary 1, $\mathbf{p}\epsilon S(a)$. Hence $\mathbf{x}\epsilon \bigcup[B(ap) : \mathbf{p}\epsilon S(a)]$. Similarly any point $\mathbf{y}\epsilon P(b)$ is contained in $\bigcup[B(qb) : \mathbf{q}\epsilon P(b)]$.

Thm 12a: If **q** bdr $P(b)$, then **qb** $= 0$, $\underrightarrow{\mathbf{qb}}$, and all successors of **b** are successors of **q**.

Proof: If **q** bdr $P(b)$, then all $N(q)$ contain a point **x** such that **xb** < 0. From axiom 5, the function $d_b^{(2)}(u)$ is continuous at point **q**, and so **qb** is not positive. Since **q** bdr $P(b)$ and $P(b)$ is an open set, it follows that **qb** is not negative. Therefore, **qb** $= 0$. From axiom 4, we know that either $\underrightarrow{\mathbf{bq}}$, or $\underrightarrow{\mathbf{qb}}$. If $\underrightarrow{\mathbf{bq}}$, then every successor of **q** is a successor of point **b** in accordance with axiom 3, theorem 1 and corollary 1. Now consider any $N_0(q)$. Theorem 9 requires that there exists a predecessor **r** of **q** and a successor **s** of **q** in $N_0(q)$. From above, **bs** < 0. In addition **rb** < 0, since the neighborhood $B(rs)$ contains **q**, and hence a point **x** such that **xb** < 0. From corollary 1, **rb** $\leqslant$ **rx** + **xb**. Therefore, $\mathbf{b}\epsilon N_0(q)$, and so all $N(q)$ contain point **b**. If **q** and **b** are the same point, then the theorem is satisfied. If they are distinct points, then the supposition $\underrightarrow{\mathbf{bq}}$, leads to a denial of axiom 1 which requires that there exists a $N(q)$ not containing **b**. Therefore, $\underrightarrow{\mathbf{qb}}$. But then since the displacement **bw** of **b** to any successor **w** is

[42] $\bigcup[A_a : \mathbf{a}\epsilon D]$ means the union of all sets A_a such that **a** is a member of D.

successor directed ($\overrightarrow{\mathbf{bw}}$), axiom 3 requires that $\mathbf{qw} \leqslant \mathbf{qb} + \mathbf{bw}$. Hence, $\mathbf{qw} < 0$ since $\mathbf{bw} < 0$, and so **w** is a successor of point **q**.

Thm 12b: If q bdr $S(a)$, then $\mathbf{aq} = 0$, $\overrightarrow{\mathbf{ay}}$, and all predecessors of **a** are predecessors of **q**.

Proof: The method of proof is the same as theorem 12*a*. Use the function $d_a^{(1)}(q)$ and axiom 5 to establish that **aq** is not positive.

Lemma 4: If some subset of a chain $C(ab)$ is identified by a property J, such that for any point **c** which has property J, all predecessors (successors) of **c** in the chain have property J, then there is either a last point in the chain having property J as one proceeds along the ordered chain in the direction of successors (predecessors).

Proof: Consider the homeomorphism of $C(ab)$ to X such that $f(a) = 0$ and $f(b) = 1$. If we agree to let all numbers in X have the property J' if their image point in $C(ab)$ has property J, then from lemma 3, it follows that for any $f(c)$ with property J', all numbers less (greater) than $f(c)$ will also have property J'. See figure 15. From the known compactness of the set X, it follows that there must be a number $f(q)$ which is either the last number in X having property J', or the first number in X not having property J' as one proceeds from 0 to 1 (1 to 0). Point **q** is therefore a member of $C(ab)$ with the required properties, and so the lemma is proved.

Thm 13A: For all $\mathbf{q} \epsilon B(ab)$, and for any $C(ab)$, there exists one and only one point $\mathbf{d} \epsilon C(ab)$ such that **q** bdr $P(d)$.

Proof: Let the property J of chain $C(ab)$ mean "successor of point **q**," then point **a** does not have property J, and point **b** does have property J. Furthermore, if a point in $C(ab)$ does have property J, then by corollary 1, all successors of that point also have property J. From lemma 4, there must either be a last point in $C(ab)$ having property J, or a first point in $C(ab)$ not having property J as one proceeds from point **b** to **a**.

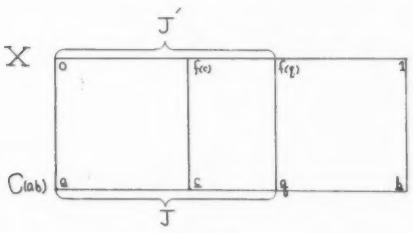


FIG. 15.

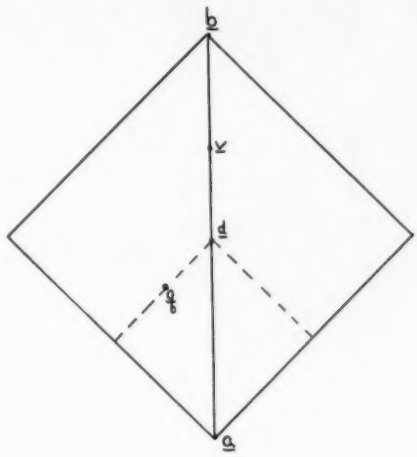


FIG. 16.

Hypothesize that there is a point **v** which is the last successor of **q** in the chain. Since by theorem 7, every $N(v)$ contains a predecessor of **v** in $C(ab)$, it follows that $B(qb)$ contains a predecessor of **v** in $C(ab)$. Hence, there is a successor of **q** in $C(ab)$ which is also a predecessor of **v** contrary to hypothesis. It follows that there must be a point **d** in $C(ab)$ which is the first non-successor of **q**; and therefore, every successor of **d** in the chain is a successor of point **q**.

From the above considerations we know that **qd** is not negative, and from theorem 7, that all $N(d)$ contain successors of **q**. Displacement **qd** is, therefore, not positive, because the function $d_q^{(1)}(u)$ must be continuous at point **d**. It follows that $\mathbf{qd} = 0$.

Consider any neighborhood $N_0(q)$ of point **q**. From theorem 9, there is a predecessor **x** of point **q** in $N_0(q)$. Since every successor **v** of point **d** in the chain $C(ab)$ is a successor of point **q**, it follows from corollary 1 that $\mathbf{xv} \leqslant \mathbf{xq} + \mathbf{qv}$. That is, for all points $\mathbf{v} \epsilon C(ab) \bigcap B(db)$, **xv** is at least as negative as **xq**. Since all $N(w)$ contain points **v** such that **xv** is at least as negative as **xq**, and since the function $d_x^{(1)}(u)$ must be continuous at point **d**, it follows that $\mathbf{xd} < 0$. Therefore $\mathbf{x} \epsilon P(d)$. Hence, all $N(q)$ contain

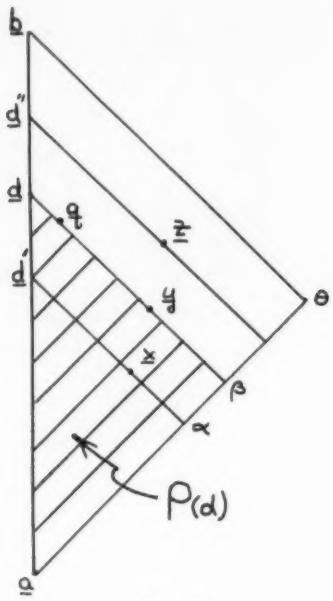


FIG. 17.

points $\mathbf{x}\epsilon P(d)$. That is, $\mathbf{q}$ lim $P(d)$. We know that $\mathbf{q}$ is not in $P(d)$ because $\mathbf{qd} = 0$. It follows that $\mathbf{q}$ bdr $P(d)$. We know furthermore that there cannot be another point $\mathbf{d}'\epsilon C(ab)$ such that $\mathbf{q}$ bdr $P(d')$, for if the displacement $\mathbf{dd}'$ is time-like, then to say that $\mathbf{qd} = 0$, $\overrightarrow{\mathbf{qd}}$, and that $\mathbf{qd}' = 0$, $\overrightarrow{\mathbf{qd}}'$ (required by thm 12*a*) would contradict axiom 3.

Thm 13b: For all $\mathbf{q}\epsilon B(ab)$, and for any $C(ab)$, there exists one and only one point $\mathbf{c}\epsilon C(ab)$ such that $\mathbf{q}$ bdr $S(c)$.

Proof: The method of proof is the same as that for theorem 13*a*.

In theorems 12*a* and 13*a*, we have shown that if a point $\mathbf{q}\epsilon B(ab)$, then relative to some chain $C(ab)$, there exists one and only one point $\mathbf{d}\epsilon C(ab)$ such that $\mathbf{q}$ bdr $P(d)$. Furthermore, it is shown that $\mathbf{qd} = 0$, and that all successors of point $\mathbf{d}$ are successors of point $\mathbf{q}$. We will use these properties for the purpose of cataloguing those members of $B(ab)$ which are not in $C(ab)$. For this purpose, the diagram in figure 17 will be helpful.

Let the hypotenuse of the right isosceles triangle in figure 17 represent chain $C_1(ad'dd''b)$, and let the area of the triangle $\mathbf{a}\theta\mathbf{b}$ represent the set $\{B(ab) - C(ab)\}$. When we speak of the area of a geometrical figure in this schematization, we mean to exclude the border lines of that figure. Hence, a member of $\{B(ab) - C(ab)\}$ is located inside of the triangle $\mathbf{a}\theta\mathbf{b}$, whereas points $\mathbf{a}$, $\mathbf{d}'$, $\mathbf{d}$, $\mathbf{d}''$ and $\mathbf{b}$ are on the line $\mathbf{a}$ to $\mathbf{b}$. The Greek letters do not represent points of sets, but only schematic markers.

We will catalogue the members of

$$\{B(ab) - C(ab)\}$$

in the following way. For all points $\mathbf{d}\epsilon C(ab)$, place boundary points of $P(d)$ on a line $\mathbf{d}\beta$ drawn parallel to $\mathbf{b}\theta$. If $\mathbf{d}'$, $\mathbf{d}$, and $\mathbf{d}''$ are three points in $C(ad'dd''b)$ with boundary points $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ respectively, then $\mathbf{z}$, $\mathbf{y}$, and $\mathbf{z}$ are placed on the lines drawn in figure 17. For any location in the triangle $\mathbf{a}\theta\mathbf{b}$, there is one and only one such line. We can now show that for any $\mathbf{d}\epsilon C(ab)$, the shaded area $\mathbf{a}\beta\mathbf{d}$ must represent the set of predecessors of $\mathbf{d}$ in $\{B(ab) - C(ab)\}$. Consider any point $\mathbf{x}$ located in $\mathbf{a}\beta\mathbf{d}$. Point $\mathbf{x}$ *bdr* $P(d')$, where $\mathbf{d}'\mathbf{d} < 0$. Hence, $\mathbf{xd} < 0$, since all successors of $\mathbf{d}'$ are successors of $\mathbf{x}$. We can furthermore show that any point $\mathbf{y}$ located in the area $\mathbf{d}\beta\theta\mathbf{b}$ is not a predecessor of $\mathbf{d}$, for if it were, then by corollary 1, $\mathbf{yd}'' < 0$

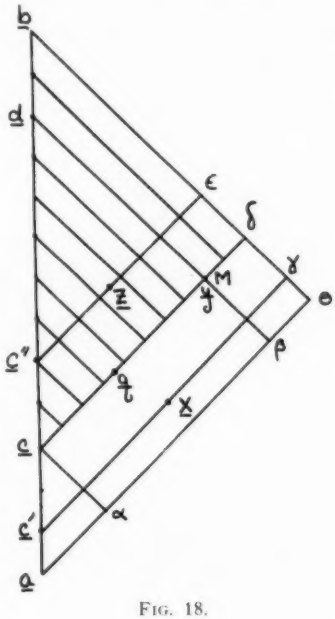


FIG. 18.

contrary to the requirement that $\mathbf{yd}'' = 0$. The diagram in figure 17 therefore portrays the relationship of members of $\{B(ab) - C(ab)\}$ to the predecessor sets of points in $C(ab)$, where those sets are the similar triangles formed by the chain points, their predecessors boundary lines, and the line $a\theta$.

In theorems 12*b* and 13*b*, we have shown that, if a point $\mathbf{q} \epsilon B(ab)$, then relative to some chain $C(ab)$, there exists one and only one point $\mathbf{c} \epsilon C(ab)$ such that $\mathbf{q}$ bdr $S(c)$. Furthermore $\mathbf{qc}$ must equal zero, and all predecessors of point $\mathbf{c}$, are predecessors of point $\mathbf{q}$. We will now use these properties to catalogue the members of $\{B(ab) - C(ab)\}$ in such a way as to extend the schematization given above.

If we first arrange the members of

$$\{B(ab) - C(ab)\}$$

relative to the predecessor sets of the points in $C(ab)$ in accordance with the above, then for some point $\mathbf{c} \epsilon C(ab)$, and for some $\mathbf{q} \epsilon \{B(ab) - C(ab)\}$ where $\mathbf{q}$ bdr $S(c)$, it is not the case that $\mathbf{q}$ will be in the area $\mathbf{a}\alpha\mathbf{c}$, or on the line $\mathbf{c}\alpha$ drawn in figure 18. We have shown that all points in $\mathbf{a}\alpha\mathbf{c}$ are predecessors of $\mathbf{c}$. Since $\mathbf{qc} = 0$, point $\mathbf{q}$ is not a predecessor of $\mathbf{c}$, and so is not in the area $\mathbf{a}\alpha\mathbf{c}$. Furthermore, if $\mathbf{q}$ bdr $P(c)$, then all $N(q)$ would contain point $\mathbf{c}$. Since the space is Hausdorff, point $\mathbf{q}$ would have to be identical with point $\mathbf{c}$ contrary to our assumption that $\mathbf{q}$ is in $\{B(ab) - C(ab)\}$. It follows that $\mathbf{q}$ is not a boundary of $P(c)$, and so will not be found on line $\mathbf{c}\alpha$. Point $\mathbf{q}$ is therefore in the area $\mathbf{c}\alpha\theta\mathbf{b}$. We will now catalogue the boundary points of $S(c)$ in the following way. Draw line $\mathbf{c}\delta$ parallel to $\mathbf{a}\theta$ (see fig. 18) and place all points in $\mathbf{c}\alpha\theta\mathbf{b}$ which are boundary points of $S(c)$ on that line. For instance, all points $\mathbf{y}$ which are boundary points of $P(d)$, and are therefore located on line $\mathbf{d}\beta$, will be moved to location M in the case they are also boundary points of $S(c)$. The procedure is then repeated for every member of $C(ab)$ by drawing a line parallel to $\mathbf{a}\theta$ from each chain member. There will be one and only one such line from any $\mathbf{c} \epsilon C(ab)$, and for any $\mathbf{v}$ in $\mathbf{c}\alpha\theta\mathbf{b}$. Clearly, this procedure may be completed without disrupting the previous arrangements of points along predecessor boundary lines.

It follows now that any point $\mathbf{z}$ in the shaded area $\mathbf{c}\delta\mathbf{b}$ is a successor of point $\mathbf{c}$; for z bdr $S(c'')$, where $\mathbf{cc}'' < 0$, hence $\mathbf{cz} < 0$ because all predecessors of $\mathbf{c}''$ are predecessors of z. Furthermore, any point $\mathbf{x}$ located in the area $\mathbf{a}\theta\delta\mathbf{c}$ is not a successor of point $\mathbf{c}$; for if it were, then by corollary 1, $\mathbf{c'x} < 0$ contrary to the requirement that $\mathbf{c'x} = 0$.

Dfn: That subset of $\{B(ab) - C(acdb)\}$ which contains only boundary points of $S(c)$ and $P(d)$ is called a *periphery* of $C(ab)$. Designate: $K(a, cd, b)$.

Each location in the triangle $\mathbf{a}\theta\mathbf{b}$ of figure 18 (such as location M), identifies a periphery set. It is easily shown that a periphery set $M = K(a, cd, b)$ can be specified independent of the particular chain through $\mathbf{c}$ and $\mathbf{d}$. We will, therefore, abbreviate the designation of a given periphery $K(m, cd, n) = K(cd)$ where $\mathbf{m}$ is understood to be any common predecessor of $\mathbf{c}$ and all members of $K(cd)$, and $\mathbf{n}$ is any common successor of $\mathbf{d}$ and all members of $K(cd)$.

Dfn: Points $\mathbf{c}$ and $\mathbf{d}$ are called the *vertices* of the periphery $K(cd)$. Point $\mathbf{c}$ is the lower *vertex*, and $\mathbf{d}$ is the upper *vertex*.

Dfn: The family of all periphery sets $K(cd)$, where $\mathbf{c}$ and $\mathbf{d}$ are included points of a chain $C_i(ab)$, is called the *peripheral family* of the chain. Designate: $\hat{\mathfrak{K}}_i(ab)$.

Dfn: The family of all periphery sets $K(cd)$, where $\mathbf{c}$ and $\mathbf{d}$ are members of a world chain C_i is called the *world peripheral family* of the world chain. Designate: $\hat{\mathfrak{K}}_i$.

Dfn: The subset of $\{B(ab) - C_i(acdb)\}$ which contains all members of $B(cd)$ is called an *umbra* of the chain $C_i(ab)$. Designate: $U_i(a, cd, b)$.

Referring again to figure 18, the umbra set $U_i(a, cd, b)$ includes all points in the triangle $\mathbf{c}M\mathbf{d}$, and is identical with $\{B(cd) - C_i(cd)\}$.

Dfn: The subset of $\{B(ab) - C_i(acdb)\}$ which is disjoint from $S(c)$ and $P(d)$ is called a *penumbra* of the chain $C_i(ab)$. Designate: $P(a, cd, b)$.

The penumbra $P(a, cd, b)$ is represented by the square $M\beta\theta\delta$ in figure 18. It follows from these definitions that every member of $K(cd)$ is also a member of the set $P(a, cd, b)$. Unlike the periphery or the umbra, the penumbra is not completely determined without specific mention of points $\mathbf{a}$ and $\mathbf{b}$.

To use successfully the constructions which have been defined, it is necessary to suppose that for all points considered, there exists a

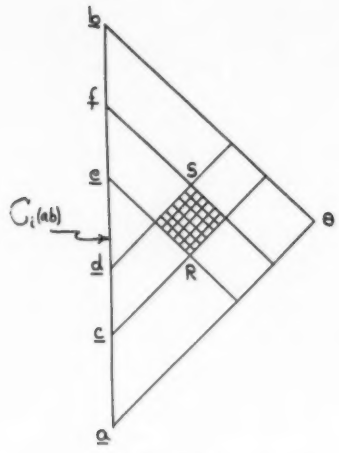


FIG. 19.

common predecessor **a** and a common successor **b**. It is not a general requirement of this axiomatization that there must be a common predecessor and a common successor of every two points: For example, there may be two chains $C_1(km)$ and $C_2(kn)$ which leave a common point **k**, and which go to points **m** and **n** such that $S(m)$ and $S(n)$ do not intersect. We do not want to discriminate against a divergent universe of this kind, but this axiomatization is not well suited to the treatment of such a universe in the large. We can, however, apply the definitions and forthcoming axioms semi-locally about each point. That is, if the universe is divergent, then interest must be restricted to the basic set system of each point. In our example, chains $C_1(km)$ and $C_2(kn)$ diverge from point **k**; but the semi-local relation between **k** and the chains is still susceptible to treatment in the above language since every $N(k)$ contains successors of **k** in both $C_1(km)$ and $C_2(kn)$.

Any two member points of a world chain C_i identify a member of the family $\hat{\mathfrak{K}}_i$, and any member of $\hat{\mathfrak{K}}_i$ identifies a pair included in C_i. We will employ the uniqueness to construct a binary function k_i on the domain of C_i and range $\hat{\mathfrak{K}}_i$. For any two subsets $V_l{}^i$ and $V_u{}^i$ of C_i, let the function k_i choose all point pairs (cd), such that $\mathbf{cd} < 0$, $\mathbf{c} \epsilon V_l{}^i$, and $\mathbf{d} \epsilon V_u{}^i$, and transform each pair into the corresponding periphery $K(cd)$. The image of the domain sets $V_l{}^i$ and $V_u{}^i$ is therefore a subfamily of $\hat{\mathfrak{K}}_i$. That is: $k_i[V_l{}^i, V_u{}^i] \equiv \hat{M} \subseteq \hat{\mathfrak{K}}_i$. Conversely, any subfamily $\hat{M}$ of $\hat{\mathfrak{K}}_i$, transforms into the set $V_l{}^i$ of all its lower vertex points, and the set $V_u{}^i$ of all its upper vertex points under the inverse function $k_i{}^{-1}$. In the case $V_l{}^i$ or $V_u{}^i$ is the set of included points of a subchain of C_i, we will designate it: $V_l{}^i(cd)$ or $V_u{}^i(cd)$, indicating that it is identical with the set $C_i \bigcap B(cd)$. From these definitions, it follows that the peripheral subfamily $\hat{U} = k_i[V_l{}^i(cd), V_u{}^i(cd)]$ includes all peripheries whose member points are in $U(a, cd, b)$.

We will now construct a topology $\mathbf{T}_2{}^{(i)}$ on the family $\hat{\mathfrak{K}}_i$. The superscript (i) of $\mathbf{T}_2{}^{(i)}$ refers to the particular world chain C_i for there are as many decompositions of points of the universe into peripheral families as there are world chains. Hence, there are equally many topologies of the kind to be constructed.

Dfn: A subfamily $\hat{Q} \subset C\hat{\mathfrak{K}}_i$ is called a *quadrifamily* if $\hat{Q} = k_i[V_l{}^i(cd), V_u{}^i(ef)]$. Designate: $\hat{Q}^i(cd, ef)$.

The shaded region of the triangle **a**ϑ**b** in figure 19 represents the quadrifamily $k_i[V_l{}^i(cd), V_u{}^i(ef)]$. Every periphery in that region has a lower vertex on the chain between **c** and **d**, and an upper vertex on the chain between **e** and **f**. The peripheries $S = K(df)$, $R = K(ce)$ are not included in the quadrifamily. More generally, no periphery $K(ij)$, where **i** or **j** is point **c**, **d**, **e** or **f**, is a member of the quadrifamily $Q(cd, ef)$.

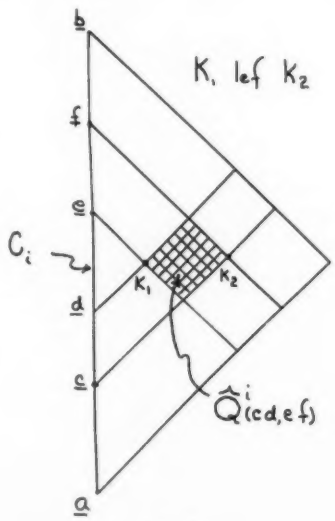


FIG. 20.

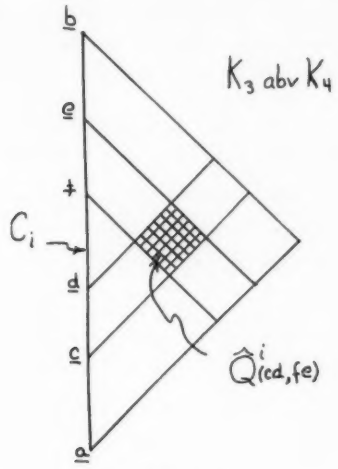


FIG. 21.

Dfn: An open family of $T_2^{(i)}$ is any subfamily of $\hat{\mathfrak{K}}_i$ which is equal to the union of a finite intersection of quadrifamilies $\hat{Q}^i$.

Clearly, the collection of all quadrifamilies is the subbase for the topology $T_2^{(i)}$.

Dfn: *A neighborhood of the periphery* K relative to $T_2^{(i)}$ is any subfamily $\hat{M}$ or $\hat{\mathfrak{K}}_i$ such that there exists an open subfamily $\hat{S} \subseteq \hat{M}$ to which K belongs. That is, if $K \epsilon \hat{S}$, $\hat{S} \epsilon T_2$, and $\hat{S} \subseteq \hat{M}$, then family $\hat{M}$ is a T_2-neighborhood of K.

Dfn: A limit periphery of a family $\hat{M} \epsilon T_2^{(i)}$ is a periphery K such that every $T_2^{(i)}$-neighborhood of K contains a member of the family $\hat{M} - [K]$.

When we say that a periphery set $K(cd)$ is null, we mean that there is no point $\mathbf{q} \epsilon K(cd)$. When we say that a peripheral family $\hat{M} \subset \hat{\mathfrak{K}}_i$ is null, we mean to deny that there exists a member of $V_l{}^i$ which is a predecessor of a member of $V_u{}^i$, where $(V_l{}^i, V_u{}^i) = k_i{}^{-1}[\hat{M}]$. From these definitions, it follows that if $Q^i(cd, ef)$ is non-null, then families $\hat{Q}^i(dc, ef)$, $\hat{Q}^i(cd, fe)$, and $\hat{Q}^i(dc, fe)$ are null. In addition, the families $\hat{Q}^i(cc, ef)$, $\hat{Q}^i(cd, ee)$, and $\hat{Q}^i(cc, ee)$ are null.

Dfn: The periphery $K_1 = K(de)$ is *to the left of* $K_2 = K(cf)$ if $\hat{Q}^i(cd, ef)$ is non-null for any C_i containing **c**, **d**, **e**, and **f**. See figure 20. Designate: K_1 lef K_2.

Dfn: The periphery $K_3 = K(de)$ is *above* $K_4 = K(cf)$ if $\hat{Q}^i(cd, fe)$ is non-null for any C_i containing **c**, **d**, **e**, and **f**. See figure 21. Designate: K_3 abv K_4.

Dfn: The periphery $K \epsilon \hat{\mathfrak{K}}_i$ is *between* peripheries $K(de)$ and $K(cf) \epsilon K_i$ if $K \epsilon [\hat{Q}^i(cd, ef) \cup \hat{Q}^i(cd, fe) \cup \hat{Q}^i(dc, ef) \cup \hat{Q}^i(dc, fe)]$.

Dfn: A subfamily $\hat{L} \subset \hat{\mathfrak{K}}_i$ is called a *horizontal peripheral line* if: (1) there are at least two members of $\hat{L}$, (2) for every two members K_1 and K_2 of $\hat{L}$, either K_1 lef K_2 or K_2 lef K_1, (3) for every two members K_1, $K_2 \epsilon \hat{L}$, there exists a periphery $K_3 \epsilon \hat{L}$ between K_1 and K_2 and (4) every limit periphery of $\hat{L}$ is a member of $\hat{L}$.

Dfn: A subfamily $\hat{L} \subset \hat{\mathfrak{K}}_i$ is called a *vertical peripheral* line if: (1) there are at least two members of $\hat{L}$, (2) for every two members K_1 and K_2 of $\hat{L}$, either K_1 abv K_2 or K_2 abv K_1, (3) for every two members K_1 and $K_2 \epsilon \hat{L}$, there exists a periphery $K_3 \epsilon \hat{L}$ between K_1 and K_2, and (4) every limit periphery of $\hat{L}$ is a member of $\hat{L}$.

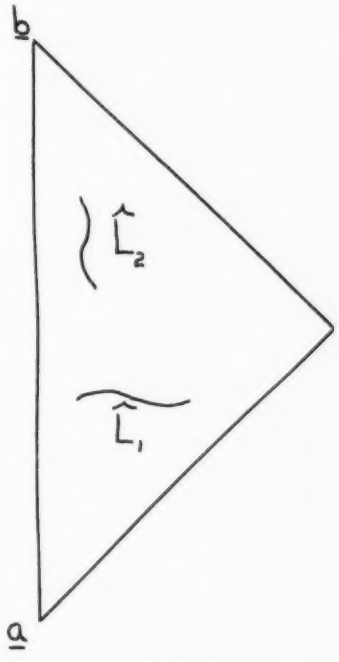


FIG. 22.

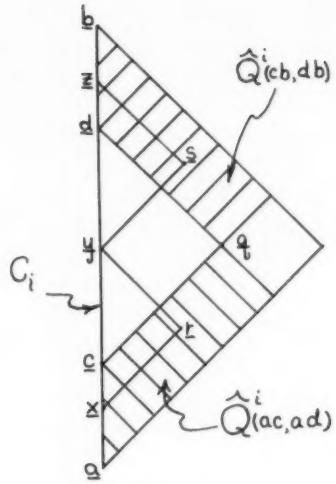


FIG. 23.

A horizontal periphery line $\hat{L}_1$ is represented in figure 22 together with a vertical peripheral line $\hat{L}_2$. In our schematization, $\hat{L}_1$ is a continuous line extending left and right such that no two members are angularly displaced from one another by less than forty-five degrees from the vertical. Line $\hat{L}_2$, on the other hand, is continuously extended up and down such that no two members are angularly displaced from one another by more than forty-five degrees from the vertical.

Thm 14: If $\mathbf{qs} < 0$, $\mathbf{q} \epsilon K_1(cd)$, and $\mathbf{s} \epsilon K_2$, where K_1, $K_2 \epsilon \hat{\mathfrak{K}}_i(ab)$, then $K_2 \epsilon \hat{Q}^i(cb, db)$.

Proof: Since $\mathbf{q}$ bdr $S(c)$, $\mathbf{cq} = 0$, $\mathbf{cq}$ from theorem 12*b*. Therefore, $\mathbf{cs} \leqslant \mathbf{cq}\ \mathbf{qs}$, and so point $\mathbf{s}$ is a successor of point $\mathbf{c}$. It follows that the lower vertex $\mathbf{v}$ of K_2 is a successor of $\mathbf{c}$. See figure 23. That is, $\mathbf{y} \epsilon C_i(ab) \bigcap B(cb) = V_l^i(cb)$. Furthermore, if $\mathbf{z}$ is the upper vertex of point $\mathbf{s}$, then $\mathbf{s}$ bdr $P(z)$, and from theorem 12*a*, $\overrightarrow{\mathbf{sz}}, \mathbf{sz} = 0$. But then $\mathbf{qz} \leqslant \mathbf{qs} + \mathbf{sz}$, so that point $\mathbf{z}$ must be a successor of $\mathbf{d}$. That is, $\mathbf{z} \epsilon C_i(ab) \bigcap B(db) = V_u^i(db)$. Periphery K_2 is therefore in $\hat{Q}^i(cb, db)$.

Thm 15: If $\mathbf{rq} < 0$, $\mathbf{q} \epsilon K_1(cd)$, and $\mathbf{r} \epsilon K_3$, where K_1, $K_3 \epsilon \hat{\mathfrak{K}}_i(ab)$, then $K_3 \epsilon \hat{Q}^i(ac, ad)$.

Proof: If $K_3 = K_3(xy)$, then we know from theorem 14 that $K_1 \epsilon \hat{Q}^i(xb, yb)$. See figure 23. It follows that $\mathbf{x}$ is a predecessor of $\mathbf{c}$, and so $\mathbf{x} \epsilon C(ab) \bigcap B(ac) = V_l^i(ac)$. Additionally point $\mathbf{y}$ is a predecessor of $\mathbf{d}$, and so $\mathbf{y} \epsilon C(ab) \bigcap B(ad) = V_u^i(ad)$. Therefore, $K_3 \epsilon \hat{Q}^i(ac, ad)$.

Thm 16: If $\mathbf{x} \epsilon B(qp)$, $\mathbf{q} \epsilon K_1(ce)$, $\mathbf{p} \epsilon K_2(df)$, $\mathbf{x} \epsilon K_3$, and K_1, $K_2 \epsilon \hat{\mathfrak{K}}_i(ab)$, then $K_3 \epsilon \hat{Q}^i(cd, ef)$.

Proof: From theorem 14, $K_3 \epsilon \hat{Q}^i(cb, eb)$ from theorem 15, $K_3 \epsilon \hat{Q}^i(ad, af)$. Therefore,

$$K_3 \epsilon \hat{Q}^i(ad, af) \bigcap \hat{Q}^i(cb, eb).$$

Since

$$V_l^i(ad) \bigcap V_l^i(cb) = V_l^i(cd),$$

and

$$V_u^i(af) \bigcap V_u^i(ef) = V_u^i(ef),$$

it follows that $K_3 \epsilon \hat{Q}^i(cd, ef)$. See figure 24.

Thm 17: If $\mathbf{q}$ and $\mathbf{p}$ are contained in the same periphery $K(cd)$, then $\mathbf{qp}$ is spacelike.

Proof: Suppose $\mathbf{qp} < 0$. Then from theorem 16, any $\mathbf{x} \epsilon B(qb)$ will be contained in the periphery $K \epsilon Q^i(cc, dd)$. But the subfamily $\hat{Q}^i(cc, dd) \subset \hat{\mathfrak{K}}_i$ is null. Therefore, $\mathbf{qp}$ cannot be negative. The same follows if $\mathbf{pq}$ is negative, and so $\mathbf{qp}$ is spacelike.

Thm 18: If $\mathbf{p}$ and $\mathbf{q}$ are members of a horizontal line in $\hat{\mathfrak{K}}_i(ab)$ then $\mathbf{pq}$ is spacelike.

Proof: If $\mathbf{p}$ and $\mathbf{q}$ are members of the same periphery, then by theorem 17, they are space-

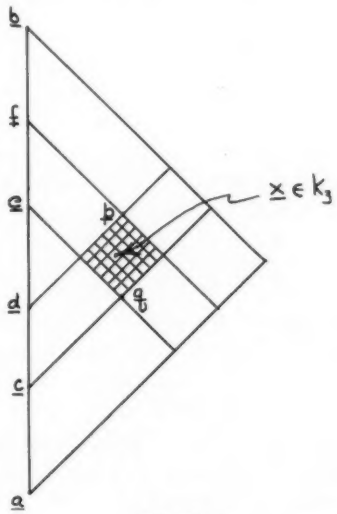


FIG. 24.

like. If $\mathbf{p}\epsilon K_1(ef)$ and $\mathbf{q}\epsilon K_2(cd)$, and if $\mathbf{qp} < 0$, then by theorem 14, $K_1\epsilon\hat{Q}^i(cb, db)$. It follows that the quadrifamily $\hat{Q}^i(ce, df)$ must be non-null. But then both $\hat{Q}^i(ce, fd)$ and $\hat{Q}^i(ec, df)$ are null contrary to the requirement that either K_1 is to the left of K_2, or K_2 is to the left of K_1. It follows that $\mathbf{qb}$ is not negative, and so the theorem is proved.

We have defined a function k_i which maps pairs of points in some world chain into the accompanying peripheries. We now wish to define a single valued function h_i on the domain of points in $\{\bigcup[B(ab): \mathbf{a}, \mathbf{b}\epsilon C_i] - C_i\}$ to the range of peripheries in $\hat{\mathfrak{K}}_i$ such that $h_i(q) = K$ where $\mathbf{q}\epsilon K\epsilon\hat{\mathfrak{K}}_i$. Let $R_i = \{\bigcup[B(ab): \mathbf{a}, \mathbf{b}\epsilon C_i] - C_i\}$. We then define the family $\hat{R}_i$ to be the family of non-null peripheries: $h_i[R_i] = \hat{R}_i \subseteq \hat{\mathfrak{K}}_i$. In case every periphery set in $\hat{\mathfrak{K}}_i$ is non-null, then $h_i[R_i] = \hat{R}_i = \hat{\mathfrak{K}}_i$.

We will frequently use the notation $R_i(ab) = R_i \bigcap B(ab) = \{B(ab) - C_i(ab)\}$. In that case $h_i[R_i(ab)] = \hat{R}_i(ab) \subseteq \hat{\mathfrak{K}}_i(ab)$.

Thm 19: If the included points of $C_1(ab)$ and $C_2(ab)$ are disjoint, then the family

$$\hat{C}_2 = h_1[C_2(ab) - \{a, b\}]$$

is a vertical peripheral line in $\hat{\mathfrak{K}}_i(ab)$.

Proof: 1. For $\mathbf{p}, \mathbf{q}\epsilon C_2(ab) - \{ab\}$, the displacement $\mathbf{pq}$ is timelike; and so, according to theorem 18, $\mathbf{p}$ and $\mathbf{q}$ are not in the same periphery. There are, therefore at least two peripheries in the image of h_1. See figure 25.

2. The displacement between any two points on the chain $C_2(ab)$ is timelike. Let $\mathbf{qp} < 0$, $\mathbf{q}\epsilon K_1(cd)$, $\mathbf{p}\epsilon K_2(ef)$. Then by theorem 14, $K_2\epsilon\hat{Q}^1(db, db)$. But then $V_l(ce)$ and $V_u(df)$ are non-null, and so $\hat{Q}^1(ce, df)$ is non-null, therefore, K_2 abv K_1.

3. For every two points $\mathbf{q}\epsilon K_1(cd)$ and $\mathbf{p}\epsilon K_2(ef)$, there exists a point $\mathbf{x}\epsilon C_2(ab)\bigcap B(qp)$. From theorem 16, $h_1(q)\epsilon\hat{Q}^1(ce, df)$. Periphery $h_1(x)$ is therefore between K_1 and K_2.

4. Let a periphery $K_0\epsilon\hat{\mathfrak{K}}_1(ab)$ be a limit periphery of the family $\hat{C}_2 = h_1[C(ab) - \{ab\}]$. It follows that no number of $\hat{C}_2$ is to the left of K_0; and that K_0 is not to the left of any member of $\hat{C}_2$. To show this, hypothesize that K_1 is to the left of K_0. It is then possible to find a quadrifamily $\hat{Q}_0^1$ which contains K_0 but does not contain K_1. If K_0 is a limit periphery of $\hat{C}_2$, then $\hat{Q}_0^1$ must include at least one other member of K_2 of $\hat{C}_2$. But then K_1 lef K_2 contrary to above proof (2). Therefore, the hypothesis is contradicted. A contradiction also follows if we assume K_0 to be to the left of K_1. It is therefore the case that for every $K_1\epsilon\hat{C}_2$, either K_1 abv K_0 or K_0 abv K_1.

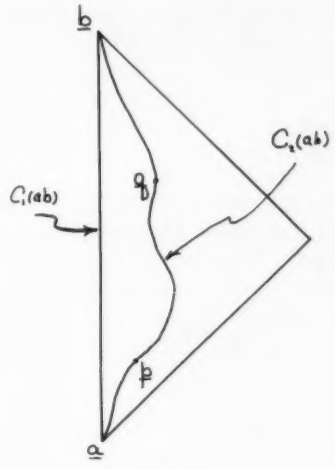


FIG. 25.

Let K_1 abv K_0, and suppose there is no member of $\hat{C}_2$ between K_0 and K_1. It is possible to find a quadrifamily $Q_0(cd, ef)$ which contains K_1 but not K_0. Hence, the neighborhood $B(cd)$ will contain the point $\mathbf{q} = h_i^{-1}[K_1]$, but no predecessor of $\mathbf{q}$ in $C_2(ab)$, contrary to theorem 7. There is, therefore, no first periphery K_1 above K_0. Similarly, there is no last periphery K_2 such that K_0 abv K_2. If we classify the members of $C_2(ab)$ according as their image peripheries are above or not above K_0, it follows from lemma 4 that there must be at least one point $\mathbf{m}\epsilon C_2(ab)$ such that $h_1(m)$ is not above K_0, and K_0 is not above $h_1(m)$. Since every point of $C_2(ab)$ maps into one and only one periphery in $\hat{K}_1(ab)$, it follows that $h(m) = K_0$, and so $K_0\epsilon\hat{C}_2$. Therefore, the family $\hat{C}_2$ contains all of its limit peripheries, and so qualifies as a vertical peripheral line.

Thm 20: If $\mathbf{q}\epsilon K_1$, $\mathbf{p}\epsilon K_2$, and K_1 lef K_2, then $\mathbf{qp}$ is positive and spacelike.

Proof: If $K_1(cd)$ and $K_2(ef)$, then both are members of some horizontal line $\hat{L}\subset\hat{\mathfrak{K}}_1$. Hence by theorem 18, $\mathbf{qb}$ is spacelike.

Consider some chain $C_2(apb)$ where point **a** is predecessor of **e** and point **b** is a successor of **f**. Let $\hat{C}_2 = h_1[C_2(apb) - \{ab\}]$, and $\hat{Q}_0^1 = \hat{Q}_0(ec, df)$. Then $h_1(p) \epsilon \hat{C}_2 \bigcap \hat{Q}_0$, and so there must be a successor **x** of **p** and a predecessor **y** of **p** in $C_2(apb)$ such that $h_1(x)$, $h_1(y) \epsilon \hat{C}_2 \bigcap \hat{Q}_0$. If **qp**, then $\mathbf{qx} \leqslant \mathbf{qp} + \mathbf{px}$ contrary to theorem 14. If $\underrightarrow{\mathbf{pq}}$, then $\mathbf{yq} \leqslant \mathbf{yp} + \mathbf{pq}$ contrary to theorem 14. It follows that **qp** is not zero, otherwise by axiom 6, $\underrightarrow{\mathbf{qp}}$, or $\underrightarrow{\mathbf{pq}}$. Therefore, $\mathbf{yp} = \mathbf{pq} > 0$.

Thm 21: If $\mathbf{m} \epsilon \{C_1(ab) - B(cd)\}$, and $\mathbf{q} \epsilon K(cd)$, where $\mathbf{c}, \mathbf{d} \epsilon C_1(ab)$. Then **mq** is positive and spacelike.

Proof: Consider the family $\hat{C}_2 \bigcap \hat{Q}_0$ where $\hat{C}_2 = h_1[C_2(aqb)]$, $\hat{Q}_0^1 = \hat{Q}^1(am, mb)$. There must be a successor **x** of **q** and a predecessor **y** of **q** in $C_2(aqb)$ such that $h_1(x)$, $h_1(y) \epsilon \hat{C}_2 \bigcap \hat{Q}_0$. If $\underrightarrow{\mathbf{mq}}$, then $\mathbf{mx} \leqslant \mathbf{mq} + \mathbf{qx}$ contrary to theorem 14. If $\underrightarrow{\mathbf{qm}}$, then $\mathbf{ym} \leqslant \mathbf{yq} + \mathbf{qm}$ contrary to theorem 14. It follows that **mq** is not negative or zero otherwise by axiom 4, $\underrightarrow{\mathbf{mq}}$ or $\underrightarrow{\mathbf{qm}}$. Therefore, $\mathbf{mq} = \mathbf{qm} > 0$.

In discussing the notion of continuous function, we gave an example of a function f on the domain of points in the two dimensional plane, and over a range in the three dimensional cube. We called the inverse f^{-1} of this function a projection of the cube into the plane. The example

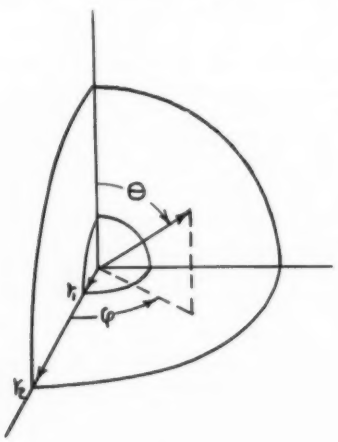


FIG. 26.

was chosen to illustrate a many valued continuous function. It is more natural in cases of this kind to use the inverse function f^{-1} directly. Refer again to figure 12.

Dfn: A function f is called a *projection* of the space $(B, \mathbf{T}_b)$ into the space $(A, \mathbf{T}_a)$ if: (1) the domain of f is B, and the range is A, (2) f is single valued and continuous, and (3) the inverse f^{-1} is continuous.

We now employ a function g on a domain of points in the set $R_i(ab) = \{B(ab) - C_1(ab)\}$ and over a range of values in an algebraic model of three dimensional Euclidean space. The subspace of particular interest in this model is the surface of a unit sphere. That subspace is denoted by $(S1S, \mathbf{T}_0^3)$.

Dfn: The family $\hat{\mathfrak{K}}_i(ab)$ is called *complete* if: (1) there exists a function g which is a projection of $[R_i(ab), \mathbf{T}_1]$ into $(S1S, \mathbf{T}_0^3)$, and (2) the function g is a homeomorphism of $(K, \mathbf{T}_1)$ into $(S1S, \mathbf{T}_0^3)$ for every $K \epsilon \hat{R}_i(ab)$.

The axioms of Group II will provide for a local four dimensional continuum,

GROUP II:

Axiom 7: For any world chain C_i, the function h_i and its inverse h_i^{-1} are continuous.

Axiom 8: For any world chain C_i, and any point $\mathbf{c} \epsilon C_i$, there exists a subchain $C_i(acb)$ such that $K_i(ab)$ is complete.

For any point **q**, choose a chain $C_1(aqb)$ through q such that $K_1(ab)$ is complete. We may now specify two parameters or coordinates which uniquely identify each $K(cd) \epsilon \hat{K}_1(ab)$. They are: r (radius) $= \mathbf{dq} - \mathbf{cq}$; t (time) $= \mathbf{dq} + \mathbf{cq}$. Since each periphery is a two dimensional surface in accordance with the definition of complete, we must find a way of assigning two coordinates to every point in each periphery. This is done by employing the function g which is a homeomorphism of each periphery to the surface of the unit sphere. The angular coordinates θ and φ may be carried into the space $R_1(ab)$, thereby providing a total of four coordinates r, t, θ, φ which uniquely identify each point.

Thm 22: The set of points in the peripheries of any horizontal line confined to a complete peripheral family, is a three dimensional subspace of U.

Proof: Consider any horizontal peripheral line $\hat{L}\epsilon\hat{\mathfrak{K}}_1(ab)$ where $\hat{\mathfrak{K}}_1(ab)$ is complete. The set of points in the peripheries of $\hat{L}$ is given by $L=h_1^{-1}(\hat{L})$. If the limit periphery of $\hat{L}$ on the left has a radius r_1 and the limit periphery of $\hat{L}$ which is not on the left has a radius r_2, then construct an algebraic model of a three dimensional spherical shell as shown in figure 26 employing parameters r, θ, and φ in the usual way. The space of this model will include only those points between r_1 and r_2, but all values of θ and φ. It is clear that there is a one-to-one correspondence between the model and points contained in peripheries of $\hat{L}$; for each periphery in $\hat{L}$ has assigned to it a unique value of r between r_1 and r_2, and each point in each periphery has a unique pair of parameters θ and φ. Let N_r be the real numbers assigned to the radius in our three-dimensional model. Then for any $\mathbf{T}_0$-open set in N_r, we may find an open family $\hat{V}$ in $\mathbf{T}_2{}^{(1)}$ such that the radius coordinate assignments in $\hat{L}\bigcap\hat{V}$ are identical with N_r. From axiom 7, the set of points $V=h_1^{-1}(\hat{V})$ is an open set in $\mathbf{T}_1$, hence the set $L\bigcap V$ is open in the subspace L. It follows that every open set N_r maps into an open set in the subspace L.

For any open in $N_{\theta\varphi}$ (where $N_{\theta\varphi}$ corresponds to the real number assignments to coordinates θ and φ on the surface of the spherical model), the function g maps into an open set W in $\mathbf{T}_1$ in accordance with axiom 8. The set $L\bigcap W$ is therefore an open set in the subspace L. Since the intersection $N_r\bigcap N_{\theta\varphi}$ is an open set in the model, and the intersection $L\bigcap V\bigcap W$ is an open set in the subspace L, it follows that the image of any open set in the model is an open set in L: for the openness of $N_r\bigcap N_{\theta\varphi}$ requires that N_r and $N_{\theta\varphi}$ be separately open.

Conversely, axioms 7 and 8 may be used to establish that any open set in L is an open set in the model. The subspace L and the model are therefore homeomorphic, and hence have the same dimension.

Thm 23: The set of points in the peripheries of any vertical line confined to a complete peripheral family, is a three-dimensional subspace of U.

Proof: The method of proof is the same as that employed in theorem 22, except that the parameter t is used to replace the parameter r in the construction of an algebraic model.

For any point $\mathbf{q}$ we may establish a local Cartesian coordinate system with point $\mathbf{q}$ at its origin and the t-axis along a chain containing $\mathbf{q}$. This is done by transforming the parameters r, t, θ, φ into x, y, z, t in the usual way. The system x, y, z, t is unduly restricted because it extends only over the points in some basic set determined by the subchain relative to which r, t, θ, φ were originally specified. It is desirable to extend the notion of a coordinate system to include some world chain C_i. The construction of such a system is as follows:

Construction 1: For any world chain C_i and some point $\mathbf{a}\epsilon C_i$, identify each periphery $K(xy)$ in the family $\hat{\mathfrak{K}}_i$ by the parameters r and t such that: r_i (radius) $=\mathbf{ya}-\mathbf{xa}$; $t_i{}^a$ (time) $=\mathbf{ya}+\mathbf{xa}$. For every time $t_i{}^a$, axiom 8 assures that there will be a radius $r_i=\mathbf{ya}-\mathbf{xa}$ such that the basic set $B(xy)$ is complete. Choose the largest such radius, or, if there is no largest, choose all peripheries at the time $t_i{}^a$, and assign coordinates θ and φ. Then perform the transformation to Cartesian coordinates x, y, z, t.

Dfn: For any world chain C_i and any point $\mathbf{a}\epsilon C_i$, a coordinate system which is erected in accordance with Construction 1 is called a *chain coordinate system* of C_i with an *origin* at point $\mathbf{a}$. The particular choice of coordinates θ and φ will be indicated by w. Designate: $O_i{}^w(a)$.

Dfn: Any point $\mathbf{p}$ which is a member of C_i or a periphery in which parameters θ and φ have been specified for the chain coordinate System $O_i{}^w(a)$ will be said to be identified by $O_i{}^w(a)$. Designate: $\mathbf{p}$ id $O_i{}^w(a)$

Dfn: The *metric function* $S_a{}^2$, relative to some point $\mathbf{a}$, is equal to $d_a{}^{(1)}(u)$ over the set of all non-predecessors of point $\mathbf{a}$, and equal to $d_a{}^{(2)}(u)$ over the set of all predecessors of point $\mathbf{a}$.

Dfn: If the metric function $S_a{}^2$ is differentiable[43] relative to a chain coordinate system $O_i{}^w(b)$, at all points $\mathbf{c}$ where $\mathbf{c}$ id $O_i{}^w(b)$, then $O_i{}^w(b)$ is called **a**-*metrical*.

Dfn: If a chain coordinate system $O_i{}^w(a)$ is **b**-metrical for every point $\mathbf{b}$ id $O_i{}^w(a)$, then $O_i{}^w(a)$, is called *metrical*.

---

[43] Differentiable will mean that first and second derivatives exist.

GROUP III:

Axiom 9: For every point **a**, there exists a world chain C_i including **a** and a choice w of θ and φ coordinates such that $O_i^w(a)$ is metrical.

From the above definitions it follows that the function S_a^2 may be expressed in a Taylor expansion about point **a** relative to a metrical chain coordinate system $O_i^w(a)$. Suppose for some point **a**, we let the system $O_i^w(a)$ also be denoted by $[klm]$, where $k, l, m = 1, 2, 3, 4$ (i.e., $x^1 = x, x^2 = y, x^3 = z, x^4 = t$). Then since the displacement from **a** to itself is zero, and since S_a^2 is a locally even function in all coordinates, it follows that the first term appearing in the Taylor expansion will be second order. Therefore, the differential form of S_a^2 at point **a** will be expressed by the quadrature $dS_a^2 = g_{kl}dx^k dx^l$, where we use the summation convention over indices which are repeated in a product.

Dfn: The quadratic array of numbers g_{kl} is called the *covariant metric tensor* at point **a** relative to the metrical chain coordinate system $O_i^w(a) = [klm]$.

Thm 24: For any metrical chain coordinate system $O_i^w(a) = [klm]$, the determinant $|g_{kl}|$ at point **a** is negative.

Proof: (1) Since the function dS_a^2 is negative for a differential displacement confined along the coordinate axis dx^4 (i.e., dt), it follows that $g_{44} < 0$. (2) The function dS_a^2 is furthermore positive for any differential displacement such that $dx^4 = 0$ (i.e., $dx^4 = 0$, $dS_a^2 \neq 0 : \supset : dS_a^2 = g_{kl}dx^k dx^l > 0$) because of the nature of the chain coordinate system that has been constructed. The necessary and sufficient condition[44] for this to be true is that all subdeterminants of the array g_{kl} are positive when $k, l = 1, 2, 3$.

$$g_{kk} > 0; \quad \begin{vmatrix} g_{kk} & g_{kl} \\ g_{lk} & g_{ll} \end{vmatrix} > 0; \quad \begin{vmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{vmatrix} > 0.$$

From this requirement together with the condition that $g_{44} < 0$, it follows that:

$$g = |g_{kl}| = \begin{vmatrix} g_{11} & g_{12} & g_{13} & g_{14} \\ g_{21} & g_{22} & g_{23} & g_{24} \\ g_{31} & g_{32} & g_{33} & g_{34} \\ g_{41} & g_{42} & g_{43} & g_{44} \end{vmatrix} < 0. \qquad k, l = 1, 2, 3, 4$$

[44] Møller, C., *op. cit.*, 235.

This may be shown by diagonalizing the subdeterminant which is the minor of g_{44}, and evaluating in the usual way.

Dfn: The *contravariant metric tensor* at points relative to the metrical chain coordinate system $O_i^w(a) = [klm]$ is the array $g^{kl} = 1/g$ (cofactor g_{kl}).

We have defined a covariant and contravariant metric tensor at a single point **a** relative to some metrical chain coordinate system $O_i^w(a)$. It is now possible to define the tensors g_{rs} at any point **b** id $O_i^w(a)$ relative to some metrical chain coordinate system $O_j^u(b) = [rst]$. Since **b** id $O_i^w(a)$, there must exist a neighborhood $N(b)$ of **b** such that for every point $\mathbf{c} \epsilon N(b)$, **c** id $O_i^w(a)$ and **c** id $O_j^w(b)$. Therefore in the neighborhood $N(b)$ the transformation from coordinates $O_i^w(a)$ and $O_j^u(b)$ is (1) one-to-one, (2) real, (3) continuous,[45] and (4) differentiable. From the transformation equations we may substitute:

$$dx^r = \frac{\partial x^r}{\partial x^k} dx^k \qquad \text{and} \qquad dx^s = \frac{\partial x^s}{\partial x^l} dx^l$$

into $dS_b^2 = g_{rs}dx^r dx^s$ yielding

$$dS_b^2 = g_{rs}\frac{\partial x^r}{\partial x^k}\frac{\partial x^s}{\partial x^l} dx^k dx^l = g_{kl}{}^{(b)} dx^k dx^l$$

where

$$g_{kl}{}^{(b)} = g_{rs}\frac{\partial x^r}{\partial x^k}\frac{\partial x^s}{\partial x^l}.$$

Dfn: The quadratic array of numbers $g_{kl}{}^{(b)}$ is called the *covariant metric tensor* at point **b** relative to the metrical chain coordinate system $O_i^w(a) = [klm]$.

GROUP IV:

Axiom 10: For any metrical chain coordinate system $O_i^w(a) = [klm]$, the function g_{kl} is continuous and differentiable.

We can now give meaning to first and second derivatives of g_{kl} at point **a**. In addition it is permissible to speak of the derivatives of g_{kl} at point **a** because it follows from the differentiability of g_{kl} that the determinate g is differentiable.

[45] Weyl, Herman, *Space, time and matter*, 85, New York, Dover, 1950.

Dfn: The *affinity* relative to the metrical chain coordinate system $O_i^w(a) = [klm]$ at point **a** is:

$$\Gamma_{kl}^{\ m} = \tfrac{1}{2}g^{mn}\left[\frac{\partial g_{kn}}{\partial x^l} + \frac{\partial g_{nl}}{\partial x^k} - \frac{\partial g_{kl}}{\partial x^n}\right].$$

Dfn: The *contracted Riemann-Christoffel tensor* relative to the metrical chain coordinate system $O_i^w(a) = [klm]$ at point **a** is:

$$R_{kl} = \Gamma_{kn}^{\ m}\Gamma_{ml}^{\ n} - \Gamma_{kl}^{\ m}\Gamma_{mn}^{\ n} + \frac{\partial}{\partial k^k}\Gamma_{lm}^{\ m} - \frac{\partial}{\partial x^m}\Gamma_{kl}^{\ m}.$$

Dfn: The *scalor Riemann-Christoffel tensor* at point **a** is: $R(a) = g^{kl}R_{kl}$.

Dfn: The *matter tensor* relative to the metrical chain coordinate system $O_i^w(a) = [klm]$ at point **a** is:

$$T_{kl} = \frac{1}{K}\left[\tfrac{1}{2}Rg_{kl} - R_{kl}\right]$$

where:

$$K = 2.073 \cdot 10^{-48}\ \frac{\text{sec}^2}{\text{gm cm}}.$$

GROUP V:

Axiom 11: For any metrical chain coordinate system $O_i^w(a) = [klm]$, if the matter tensor T_{kl} is non-zero, then T_{44} is positive.

It is now possible to construct a local system of inertia at any point **a**. This is done in the following way:

Construction **2:** For any metrical chain coordinate system $O_i^w(a) = [klm]$ at point **a**, transform to coordinate system $[\mu\nu\sigma]$ by means of the following transformation:

$$x^\mu = \tfrac{1}{4}\delta_k^{\ \mu}\{x^k + \tfrac{1}{2}x^l x^m \Gamma_{lm}^{\ k}(a)\}$$

where: $\Gamma_{lm}^{\ k}(a)$ is a number equal to $\Gamma_{lm}^{\ k}$ at point **a**. That is, derivatives of $\Gamma_{lm}^{\ k}(a)$ are zero whereas derivatives of $\Gamma_{lm}^{\ k}$ are not in general zero. From these transformations, it follows[58] that $\dfrac{\partial x^k}{\partial x^\nu} = \delta_\nu^{\ k}$ at point **a**, and that the affinity $\Gamma_{\mu\nu}^{\ \sigma}(a) = 0$. From the definition of affinity: $\partial g_{\mu\nu}/\partial x^\sigma = \Gamma_{\mu,\nu\sigma} + \Gamma_{\nu,\mu\sigma}$ (where $\Gamma_{\mu,\nu\sigma} = g_{\mu\tau}\Gamma_{\nu\sigma}^{\ \tau}$), and so it follows that $\partial g_{\mu\nu}/\partial x^\sigma = 0$ in the coordinate system $[\mu\nu\sigma]$. From the nature of the transformation it is clear that the array $g_{\mu\nu} = \dfrac{\partial x^k}{\partial x^\mu}\dfrac{\partial x^l}{\partial x^\nu}g_{kl}$ is numerically identical with the array g_{kl}. The only effect of the transformation is to assure that the first derivatives of the metric tensor vanish. Therefore, it is still the case that the determinant $|g_{\mu\nu}| < 0$.

It will now be possible to diagonalize the metric tensor so that it has the form $g_{11} = g_{22} = g_{33} = -g_{44} = 1$; $g_{ij} = 0$ for $i \neq j$. This can be done by a transformation of the form:

$$\begin{aligned} x^\mu &= x^\mu(\mathring{x}^1, \mathring{x}^2, \mathring{x}^3) \qquad \text{where:} \quad \mu = 1, 2, 3 \\ x^\nu &= x^\nu(\mathring{x}^1, \mathring{x}^2, \mathring{x}^3, \mathring{x}^4) \qquad\quad \nu = 4 \end{aligned}$$

such that the partial derivative $\partial\mathring{x}^4/\partial x^4$ is positive.[46] The construction is complete.

Dfn: Any coordinate system connected to a metrical chain coordinate system $O_i^w(a)$ by the transformations in Construction 2 will be called a *local inertial system* at point **a**. Designate: $I(a)$.

The reason for our precaution concerning the sign of $\partial\mathring{x}^4/\partial x^4$ in the last transformation under Construction 2 is to insure that the time direction is not changed under transformation to the inertial frame. It is not enough to know that the final form of the invariant must be: $ds^2 = dx^2 + dy^2 + dz^2 - dt^2$; for the function ds^2 is insensitive to time reversal.[47]

Thm 25: For every $I(a) = [ij]$, if the matter tensor T_{ij} is non-zero, then T_{44} is positive.

Proof: By axiom 11, the above proposition holds for any metrical system $O_i^w(a)$. Since every $I(a)$ is defined by reference to some metrical $O_i^w(a) = [klm]$, then we need only consider the effect of Construction 2 on the positive T_{44} component of a non-zero tensor T_{kl}. The first transformation

$$x^\mu = \tfrac{1}{4}\delta_k^{\ \mu}\{x^k + \tfrac{1}{2}x^l x^m \Gamma_{lm}^{\ k}(a)\}$$

is such that $T_{kl} = T_{\mu\nu}$ for all $k = \mu$ and $l = \nu$. Therefore the T_{44} component of $T_{\mu\nu}$ remains positive and equal to the T_{44} component of T_{kl}. If we refer to the coordinate system $I(a) = [ij] = [\mathring{x}]$, then the second transformation:

$$\begin{aligned} x^\mu &= x^\mu(\mathring{x}^1, \mathring{x}^2, \mathring{x}^3) \qquad \text{where:} \quad \mu = 1, 2, 3 \\ x^\nu &= x^\nu(\mathring{x}^1, \mathring{x}^2, \mathring{x}^3, \mathring{x}^4) \qquad\quad \nu = 4 \end{aligned}$$

to the system $[\mathring{x}]$ has the property that $\partial x^\mu/\partial\mathring{x}^4 = 0$ where $\mu = 1, 2, 3$. Since

$$\mathring{T}_{44} = \frac{\partial x^\mu}{\partial\mathring{x}^4}\frac{\partial x^\nu}{\partial\mathring{x}^4}T_{\mu\nu} = \left(\frac{\partial x^4}{\partial\mathring{x}^4}\right)^2 T_{44},$$

it follows that $\mathring{T}_{44}$ is also positive.

---

[46] Møller, C., *op. cit.*, 274–275.
[47] Schrödinger, E., *op. cit.*, 77.

MEMOIRS OF THE
AMERICAN PHILOSOPHICAL SOCIETY

William Shippen, Jr., Pioneer in American Medical Education. BETSY COPPING CORNER.
Vol. 28. xiii, 161 pp., 9 pls. 1951. $2.75.

West Roman Vulgar Law. ERNST LEVY.
Vol. 29. xix, 305 pp. 1951. $5.00.

Letters of Benjamin Rush. L. H. BUTTERFIELD.
Vol. 30. lxxxvii, 1295 pp., 18 figs. 2 v. 1951. $15.00.

Erwin Frink Smith—A Story of North American Plant Pathology. ANDREW DENNY RODGERS III.
Vol. 31. x, 675 pp., 4 figs. 1952. $5.00.

The Primitive City of Timbuctoo. HORACE MINER.
Vol. 32. xx, 297 pp., 27 figs. 1953. $5.00.

Age and Achievement. HARVEY C. LEHMAN.
Vol. 33. xi, 359 pp., 170 figs. 1953. $7.50.

Michael Servetus—A Translation of his Geographical, Medical and Astrological Writings. CHARLES DONALD O'MALLEY.
Vol. 34. 208 pp., 6 figs. 1954. $3.00.

Thomas Jefferson's Farm Book. EDWIN MORRIS BETTS.
Vol. 35. xxiii, 722 pp., 8 pls. 1953. $12.00.

The Long Parliament, 1640–1641. A Biographical Study of its Members. MARY FREAR KEELER.
Vol. 36. ix, 410 pp. 1954. $6.00.

Astrology in Roman Law and Politics. FREDERICK H. CRAMER.
Vol. 37. x, 291 pp., 22 figs., 1 map. 1954. $5.00.

Cumulative Record of Exhibition Catalogues. The Pennsylvania Academy of the Fine Arts, 1807–1870. ANNA WELLS RUTLEDGE.
Vol. 38. v, 450 pp. 1955. $6.00.

Historic Germantown. From the Founding to the Early Part of the Nineteenth Century. A Survey of the German Township. HARRY M. and MARGARET B. TINKCOM, GRANT M. SIMON.
Vol. 39. vii, 154 pp., 241 figs., 5 maps. $5.00.

The Slave Systems of Greek and Roman Antiquity. WILLIAM L. WESTERMANN.
Vol. 40. xii, 180 pp. 1955. $3.00.

Edmund Burke, New York Agent, with his Letters to the New York Assembly and Intimate Correspondence with Charles O'Hara, 1761–1776. ROSS J. S. HOFFMAN.
Vol. 41. xiv, 632 pp., 3 figs. 1956. $6.50.

An Historian's World. Selections from the Correspondence of John Franklin Jameson. ELIZABETH DONNAN and LEO F. STOCK.
Vol. 42. xi, 382 pp., 1 fig. 1956. $6.00.

Franklin and Newton. I. BERNARD COHEN.
Vol. 43. xxvi, 657 pp., 24 figs. 1956. $6.00.

A History of Luminescence from the Earliest Times until 1900. E. NEWTON HARVEY.
Vol. 44. xxiii, 692 pp., 50 figs. 1957. $6.00.

Population Redistribution and Economic Growth, United States, 1870–1950. I. Methodological Considerations and Reference Tables. EVERETT S. LEE, ANN RATNER MILLER, CAROL P. BRAINERD, RICHARD A. EASTERLIN. Prepared under the direction of SIMON KUZNETS and DOROTHY SWAINE THOMAS.
Vol. 45. 759 pp. 1957. $5.00.

Give Me Liberty. The Struggle for Self-Government in Virginia. THOMAS J. WERTENBAKER.
Vol. 46. ix, 275 pp., 7 figs. 1958. $3.00.

Benjamin Franklin and Italy. ANTONIO PACE.
Vol. 47. xiii, 450 pp., 9 figs. 1958. $5.00.

TRANSACTIONS

OF THE

AMERICAN PHILOSOPHICAL SOCIETY

A Survey of Islamic Astronomical Tables. E. S. KENNEDY.
Vol. 46, pt. 2, 55 pp., 8 figs., 1956. $1.50.

The Interpretation of Dreams in the Ancient Near East. With a Translation of an Assyrian Dream-Book. A. LEO OPPENHEIM.
Vol. 46, pt. 3, 195 pp., 6 figs., 13 pls., 1956. $2.50.

Klamath Prehistory: The Prehistory of the Culture of the Klamath Lake Area, Oregon. L. S. CRESSMAN.
Vol. 46, pt. 4, 139 pp., 73 figs., 1956. $2.00.

Excavations at Chupícuaro, Guanajuato, Mexico. MURIEL NOÉ PORTER.
Vol. 46, pt. 5, 123 pp., 30 figs., 3 maps, 10 plans, 1956. $1.75.

The Monguors of the Kansu-Tibetan Frontier. Part II. Their Religious Life. LOUIS M. J. SCHRAM.
Vol. 47, pt. 1, 164 pp., 1957. $3.00.

Comparative Studies of North American Indians. HAROLD E. DRIVER and WILLIAM C. MASSEY.
Vol. 47, pt. 2, 292 pp., 164 maps, 31 figs., 1957. $5.00.

The Cultivation and Weaving of Cotton in the Prehistoric Southwestern United States. KATE PECK KENT.
Vol. 47, pt. 3, 276 pp., 143 figs., 13 maps, 1957. $4.00.

The Sixteenth-Century Pokom-Maya: A Documentary Analysis of Social Structure and Archaeological Setting. S. W. MILES.
Vol. 47, pt. 4, 49 pp., 4 figs., 1957. $2.00.

Maistre Nicole Oresme: Le Livre de Yconomique d'Aristote. ALBERT DOUGLAS MENUT.
Vol. 47, pt. 5, 71 pp., 1 fig., 1957. $2.00.

Nikolaos Mesarites: Description of the Church of the Holy Apostles at Constantinople. GLANVILLE DOWNEY.
Vol. 47, pt. 6, 71 pp., 1957. $2.00.

The Mycenae Tablets II. EMMETT L. BENNETT, JR. With an introduction by Alan J. B. Wace and Elizabeth B. Wace. Translations and commentary by John Chadwick.
Vol. 48, pt. 1, 122 pp., 111 figs., 1958. $3.00.

The Neolithic Age in Eastern Siberia. HENRY N. MICHAEL.
Vol. 48, pt. 2, 108 pp., 100 figs., 7 maps, 1958. $2.50.

Late Ancient and Medieval Population. J. C. RUSSELL.
Vol. 48, pt. 3, 152 pp., 1958. $4.00.

The Search for a New Voltaire. Studies in Voltaire Based upon Material Deposited at the American Philosophical Society. IRA O. WADE.
Vol. 48, pt. 4, 206 pp., 1958. $5.00.

Surface Survey of the Northern Valley of Mexico: the Classic and Post-Classic Periods. PAUL TOLSTOY.
Vol. 48, pt. 5, 101 pp., 24 figs., 1958. $2.00.

Preliminary Archaeological Investigations in the Sierra de Tamaulipas, Mexico. RICHARD S. MACNEISH.
Vol. 48, pt. 6, 210 pp., 51 figs., 1958. $5.00.

The Spectrum of Beta Lyrae. J. SAHADE, S.-S. HUANG, O. STRUVE, and V. ZEBERGS.
Vol. 49, pt. 1, 64 pp., 20 figs., 30 pls., 1959. $2.00.

Edward T. Cardwell: Peelite. ARVEL B. ERICKSON.
Vol. 49, pt. 2, 107 pp., 1959. $2.00.

Aleut Dialects of Atka and Attu. KNUT BERGSLAND.
Vol. 49, pt. 3, 128 pp., 23 figs., 1959. $3.00.